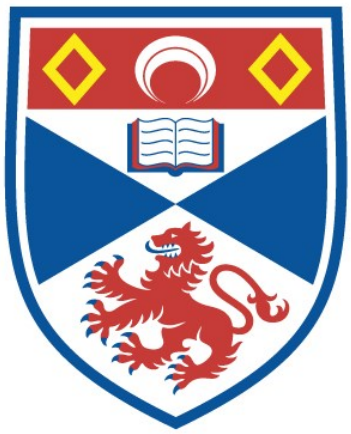
CO-CREATING DATA PROTECTION SOLUTIONS THROUGH A COMMONS

Janis Chin Ching Wong

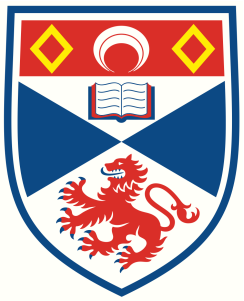
A Thesis Submitted for the Degree of PhD
at the
University of St Andrews



2022

Co-creating Data Protection Solutions Through A Commons

Janis Chin Ching Wong



University of
St Andrews

This thesis is submitted in partial fulfilment for the degree of
Doctor of Philosophy (PhD)
at the University of St Andrews

February 2022

ABSTRACT

In our data-driven society, personal data affecting individuals as data subjects is increasingly being collected and processed by sizeable, international companies. While data protection laws and privacy technologies attempt to limit the impact of data breaches and privacy scandals, they rely on individuals having a detailed understanding of the available recourse, resulting in the responsibilisation of data protection. Existing data stewardship frameworks incorporate data protection considerations and employ data-protection-by-design principles but may not include data subjects in the process itself, relying on supplementary legal doctrines to strengthen data protection enforcement. Current data protection solutions also lack support for protecting individual autonomy over personal data through co-creation and participation, particularly where there is socio-technical and communal value to collaborative data from which data subjects may not currently benefit.

These challenges motivate the application of a theoretical and practical framework that can encourage co-creation of data protection solutions, increase awareness of different stakeholder interests, and rebalance power between data subjects and data controllers. In this thesis, we propose adapting the commons framework to create a data protection-focused data commons. We conduct interviews with commons experts to identify the institutional barriers to creating a commons and challenges of incorporating data protection principles into a commons. We propose requirements for establishing a data protection-focused data commons by applying our interview findings and data protection principles. We then deploy the data protection-focused data commons using an online learning use case. We conduct a study to explore the usefulness of the commons for supporting students' agency and co-creating data protection solutions in response to tutorial recordings, their consent preferences, and attitudes towards privacy and online learning. We find that a data protection-focused data commons as a socio-technical framework can support the collaboration and co-creation of data protection solutions for the benefit of data subjects.

For the real question is whether the brighter future is really always so distant. What if, on the contrary, it has been here for a long time already, and only our own blindness and weakness has prevented us from seeing it around us and within us, and kept us from developing it?

Vaclav Havel, The Power of the Powerless

Acknowledgements

The world looks very different now in 2022 compared to when I started my PhD in 2018. I am grateful to be surrounded by supportive people who have made this journey within academia and life outwith academia enjoyable.

First, I would like to thank my supervisor, Tristan Henderson, for his words of wisdom and trust. His advice has always been honest, kind, and thoughtful. Even during my Master's, he was always telling me to give it a go; Many of my research endeavours have been fruitful because of his encouragement. My thanks also goes to my second supervisor, Kirstie Ball. At the beginning of my PhD, Kirstie encouraged me to research a topic that "sets you on fire", which was how I landed here. I am thankful for her feedback and perspective. I would also like to thank my examiners Dharini Balasubramaniam and Alison Powell for their critical questions and thoughtful insights that helped make this thesis more robust and valuable for wider audiences.

Thank you to the School of Computer Science for welcoming me since day one. At St Andrews, I thank the Centre for Research into Information, Surveillance and Privacy (CRISP) and the St Andrews Computer Human Interaction (SACHI) group for being my academic homes. Thank you to the Computer Science Admin and Systems Teams for their support and friendly conversations in equal measure. My thanks also goes to the David Russell Apartments facilities staff for looking after me.

In my broader academic life, I would like to express my gratitude to the researchers and students who I worked alongside, collaborated with, taught, and learnt from. Thank you to those who agreed to be interviewed as part of my research and to the students who took part in my study. The pandemic made participation more challenging and I greatly appreciate their time and effort. Interdisciplinary work in this field is exciting but can be difficult to navigate within traditional research institutions. I am grateful to the multidisciplinary researchers, many of whom are referenced in my thesis, who have paved the way and made this space more welcoming. In particular, I would like to thank Angela Daly for inviting me to share my research early on in Hong Kong and Scotland. To Adriana Wilde for finding opportunities for me and for her friendship. To Hamed Haddadi for the canoe polo and getting me out of my comfort zone. To Marwan Fayed for laughing at my jokes. I am glad to have gotten to know Anuj Puri, Bernard Keenan, Bran Knowles, Caroline Stockman, Catherine Stihler, Chris Jefferson, Emma Nottingham, Guido Noto La Diega, Ian Gent, Ishbel Duncan, Judith

Rauhofer, Karen Gregory, Karen Renaud, Katie Nolan, Lea Racine, Lilian Edwards, Loraine Clarke, Mahesh Marina, Mario Moreno Rocha, Meghan McNamara, Miguel Nacenta, Orla Lynskey, Rachel Marsh, Roger Von Laufenberg, Rossana Ducato, Saleem Bhatti, Tommaso Fia, Uta Hinrichs, Virginia Eubanks, and Xu Zhu. I especially want to thank everyone who double- and triple-checked that I was given due credit for my work.

During my PhD, I have been fortunate to receive opportunities to pursue alternative avenues of research. For these endeavours, I would like to thank Milly Zimeta, Olivier Thereaux, and Jeni Tennison at the Open Data Institute for supporting my fellowship as well as Sue Chadwick for her camaraderie. To the Berkman Klein Center for Internet & Society at Harvard University for inviting me to participate in their research sprints. To Talat Yaqoob and everyone in the 2020/21 cohort for creating Pass The Mic and collectively diversifying Scottish media by sharing the country's rich histories and stories. I would also like to thank the wonderful people at the Open Rights Group, St Leonards School, Robogals, and Young Women Lead.

Working on this thesis was as much of a scientific pursuit as a creative one. *Sláinte* to St Andrews and the Scottish trails. Long, summer evening walks by the sea have helped me relax and find clarity. This thesis would not have been written without the thousands of hours of music consumed. Thank you to the artists I listened to for translating emotions into audio. My thanks also go to journalists who fearlessly pursue the truth behind the stories, writing on technology and beyond. Thank you to those who engaged in discussions with me on Twitter, where I spent countless hours in conversation on the pretence of research. An ode to 香港, Hong Kong, where I spent a few months during the pandemic working on this thesis, for being the first commons, community, and commune I know.

A special mention is needed for those who managed to see me through from beginning to end. I am very lucky to have shared an office with Donald Robertson and Ryo Yanagida. Thank you for the memes, omnishambles, and putting up with my antics. Without their friendship, this PhD would have been much harder. Thank you to Sohni Chakrabarti for having my back. To Amy Tsang for cheering me on. To Pranav Putcha for travelling the distance.

Last but not least, thank you to my parents for their unwavering love and support. Amongst many things, I have them to thank for my curiosity. This thesis is for my 嫲嫲, paternal grandmother, and 婆婆, late maternal grandmother. 多謝。Without their sacrifices, none of this would have been possible.

Declaration

Candidate's declaration

I, Janis Chin Ching Wong, do hereby certify that this thesis, submitted for the degree of PhD, which is approximately 48,000 words in length, has been written by me, and that it is the record of work carried out by me, or principally by myself in collaboration with others as acknowledged, and that it has not been submitted in any previous application for any degree. I confirm that any appendices included in my thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

I was admitted as a research student at the University of St Andrews in September 2018.

I received funding from an organisation or institution and have acknowledged the funder(s) in the full text of my thesis.

10th August 2022

Date Signature of candidate

Supervisor's declaration

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree. I confirm that any appendices included in the thesis contain only material permitted by the 'Assessment of Postgraduate Research Students' policy.

10 August 2022

Date Signature of supervisor

Permission for Publication

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand, unless exempt by an award of an embargo as requested below, that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that this thesis will be electronically accessible for personal or research use and that the library has the right to migrate this thesis into new electronic forms as required to ensure continued access to the thesis.

I, Janis Chin Ching Wong, confirm that my thesis does not contain any third-party material that requires copyright clearance.

The following is an agreed request by candidate and supervisor regarding the publication of this thesis:

Printed copy

No embargo on print copy.

Electronic copy

No embargo on electronic copy.

10th August 2022

Date Signature of candidate

10 August 2022

Date Signature of supervisor

Underpinning Research Data or Digital Outputs

Candidate's declaration

I, Janis Chin Ching Wong, hereby certify that no requirements to deposit original research data or digital outputs apply to this thesis and that, where appropriate, secondary data used have been referenced in the full text of my thesis.

10th August 2022

Date Signature of candidate

FUNDING

This work was supported by the University of St Andrews St Leonards Interdisciplinary Scholarship in collaboration between St Leonards College, School of Computer Science, and School of Management.

PUBLICATIONS

This thesis is entirely my work, but has been supported by a number of collaborators. Throughout this thesis, I use the word "we" to acknowledge the contribution these collaborators have made to this work.

During the course of my PhD, I have contributed to the following publications. Where I am first author, I have been chiefly responsible for the core contributions of experimental design, implementation, and execution of studies and analyses, and it is this work that I present in this thesis.

Articles in Peer-Reviewed Conferences and Journals

- Janis Wong, Tristan Henderson, and Kirstie Ball. "Data protection for the common good: Developing a framework for a data protection-focused data commons". In: *Data and Policy (2022)* doi: 10.1017/dap.2021.40.
- Janis Wong. "Data Governance for Online Learning". In: *Open Data Institute (2021)*. **https://theodi.org/article/data-governance-online-learning/**
- Janis Wong, Tristan Henderson, and Kirstie Ball. "Data protection for the common good: Developing a framework for a data protection-focused data commons". In: *Proceedings of the 5th International Data for Policy Conference (2020)*. doi: 10.5281/zenodo.3965670
- Janis Wong. "The 'personal' in personal data: Who is responsible for our data and how do we get it back?". In: *Legal Information Management (2020)*. doi: 10.1017/S1472669620000249
- Janis Wong and Tristan Henderson. "Co-creating autonomy: Group data protection and individual self-determination within a data commons". In: *Proceedings of the 15th International Data Curation Conference (2020)*. doi: 10.2218/ijdc.v15i1.714
- Janis Wong and Tristan Henderson. "The right to data portability in practice: exploring the implications of the technologically neutral GDPR". In: *International Data Privacy Law (2019)*. doi: 10.1093/idpl/ipz008/5529345
- Janis Wong and Tristan Henderson. "How portable is portable? Exercising the GDPR's right to data portability." In: *Proceedings of the 2018 ACM*

International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers. doi: 10.1145/3267305.3274152

Articles Under Review for Conference Proceedings and Journals

- Janis Wong, Lea Racine, Tristan Henderson, and Kirstie Ball. "Online learning as a commons: Supporting students' online learning data protection preferences through a collaborative digital environment". In: *Journal of Intellectual Property, Information Technology and Electronic Commerce Law (2022)*

CONTENTS

List of Figures

List of Tables

Acronyms

ACM Association for Computing Machinery

AdTech advertising technologies

API application programming interface

ARDC Australian Research Data Commons

BMA British Medical Association

CCPA California Consumer Privacy Act

CPR common-pool resource

CQC Care Quality Commission

DPA Data Protection Authority

DPbD Data Protection by Design

DPD Data Protection Directive

DPO Data Protection Officer

EdTech education technologies

EOSC European Open Science Cloud

EU European Union

GDC Genomic Data Commons

GDPR General Data Protection Regulation

GP General Practitioner

GPDPR General Practice Data for Planning and Research

HDI human-data interaction

HEIs higher education institutions

IAD Institutional Analysis and Development

ICO Information Commissioner's Office

IoT Internet of Things

IT Information Technology

IUIPC Internet Users' Internet Privacy Concerns

JISC Joint Information Systems Committee

NDG National Data Guardian

NHS National Health Service

ODI Open Data Institute

OfS Office for Students

PETs privacy-enhancing technologies

RCGP Royal College of General Practitioners

RoA Right of Access

RtbF Right to be Forgotten

RtDP Right to Data Portability

STS Science and Technology Studies

UK United Kingdom

US United States

GLOSSARY

application programming interface A set of routines, protocols, and tools for building software applications.

big data Large amounts of data that are too large or complex to be dealt with by traditional data-processing application software and require different methods of analysis to systematically extract information from.

common-pool resource A type of good consisting of a natural or human-made resource system whose size or characteristics makes it costly, but not impossible, to exclude potential beneficiaries from obtaining benefits from and may be over-exploited.

contextual integrity The concept where adequate protection for privacy is considered alongside norms of specific contexts.

dark patterns User interface features that are designed to intentionally manipulate, steer, or nudge users into actions that may not be in their best interest.

data controller "The natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data." (GDPR Article 4(7)).

data governance A system governing the norms, principles, and rules surrounding various types of data. Data stewardship and data management are regarded as subsets of data governance.

data management The technical management of data such as the creation of digital systems and infrastructures.

data processor "A natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller." (GDPR Article 4(8)).

data protection The fair and proper use of information about people as is part of the fundamental right to privacy.

data protection-focused data commons A theoretical and practical framework that can be used to encourage co-creation of data protection solutions, increase awareness of different stakeholder interests, and rebalance power between data subjects and data controllers.

data steward Facilitates collaboration to unlock the value of data, protects actors from harms caused by data sharing, and monitors users to ensure that their data use is appropriate and can generate data insights.

data stewardship The process by which individuals or teams within data-holding organisations are empowered to actively initiate, facilitate, and coordinate data towards the public interest. Examples of data stewardship frameworks include data trusts, data foundations, and data cooperatives.

data subject An identified or identifiable natural person "who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person." (GDPR Article 4(1)).

data subject rights Rights that data subjects have under the GDPR. These are be summarised as: The right to be informed (Article 14), the right of access (Article 15), the right to rectification (Article 16), the right to be forgotten (Article 17), the right to restrict processing (Article 18), the right to data portability (Article 20), the right to object (Article 21), and rights in relation to automated decision making and profiling (Article 22).

Facebook As of 2022, the world's largest social networking site that was founded in the US in 2004.

Institutional Analysis and Development A framework developed by Elinor Ostrom that supports the creation of a commons and analyses the dynamic situations where individuals develop new norms, rules, and physical technologies.

Internet of Things Physical objects that are embedded with sensors, processing ability, software, and other technologies that connect and exchange data with other devices over the Internet or other communications networks.

Microsoft Teams Owned by Microsoft, an American multinational technology company, Microsoft Teams is a business communication platform used by organisations to facilitate communication, video conferencing, and file sharing services.

personal data "Any information relating to an identified or identifiable natural person." (GDPR Article 4(1)).

polycentricity A complex form of governance with multiple centres of decision-making, each of which operates with some degree of autonomy.

privacy A concept that broadly encompasses freedom of thought, control over one's body, solitude in one's home, control over personal information, freedom from surveillance, protection of one's reputation, and protection from searches and interrogations.

privacy-enhancing technologies A disruptive set of technologies and approaches which, when combined with changes in wider policy and business frameworks, could enable the sharing and use of data in a privacy-preserving manner. They also have the potential to reshape the data economy and to change the trust relationships between citizens, governments and companies.

responsibilisation The responsibilisation of personal data refers to the process where individuals have the burden of protecting their own personal data as opposed to it being data controllers' responsibility.

social networking site An online platform which people use to build social relationships and networks with other people who share similar personal or career content, interests, activities, backgrounds, or real-life connections. Examples of social networking sites include Facebook, Twitter, LinkedIn, Reddit, and Tumblr.

surveillance The monitoring of behaviour, activities, or information for the purpose of information gathering, influencing, managing, or directing. Surveillance can be done in person, digitally, manually, or automatically.

Twitter One of world's largest social networking sites known for its short form posts (tweets) that was founded in the US in 2006.

Chapter One

Introduction

Rapid technological innovation in the past decade has changed how we, as individuals, interact with companies using our personal data. Personal data refers to data that is collected, stored, analysed, and processed about us as individuals. As technology becomes more advanced, the widespread use of computers, smart phones, and Internet of Things (IoT) devices has magnified the use of technology in our everyday lives, collecting more and more data about us and pushing us towards a "data-driven society" [182]. These large datasets, or big data, have led to surveillance networks [13], surveillance capitalism [273], and the globalisation of surveillance industries [12] becoming the norm. In a dataveillance environment, our data and metadata are constantly being collected, even during our sleep [38]. With the greater collection, sharing, and analysis of data, it is important to consider how our personal data could be protected both as individuals and as a collective.

Personal data is increasingly, knowingly or unknowingly, being gathered about us in an attempt to learn more about our behaviours, patterns, and consumer preferences in the digital environment. In the private sphere, the omnipresence of IoT devices track all individuals' movements within the spaces they occupy. Data is sent back and forth between homes and servers, creating new streams of information about the owners' daily lives that could be used for further profiling and targeting [117]. Our shopping basket [115], search results [181], music playlists [150], and travel patterns [2] all detail the intricacies of our decision-making, where such data is gathered and used to serve us adverts based on our habits [195]. In the public sphere, the companies that gather citizens' data to process and analyse are increasing every day. For example, Street View, a service offered by Google Maps is a "virtual representation of our surroundings on Google

Maps, consisting of millions of panoramic images" that have either been gathered by Google or offered by contributors [94]. While the service allows anyone from anywhere in the world to see the street view of public places, vulnerable groups such as abortion protesters, sunbathers, nose-pickers, and men leaving strip clubs, can be captured and exposed online without their knowledge or consent [142]. In public healthcare, governments are also increasingly digitising our health data with the aim of improving healthcare delivery [67] and for healthcare research [128]. The Facebook and Cambridge Analytica case also demonstrates how a vast amount of people's personal data can be improperly shared [208], where profile data is analysed for generating data analytics that has the potential to influence democracy [31]. With these forms of personal data being increasingly collected and stored in our data-driven society, data breaches and privacy scandals have also come to light with greater frequency [29]. As a result, data protection and privacy have come to the forefront, where more people are more cautious about what information they put online in an attempt for individuals to take back control over their personal data.

To address some of these concerns, new laws such as the European General Data Protection Regulation (GDPR) [73] and California Consumer Privacy Act (CCPA) [32] have been implemented in an attempt to rebalance power between citizens and the increasingly sizeable and international companies that are collecting, and potentially exploiting, their data. New projects have also attempted to give users the ability to manage their own data through technological tools and platforms, such as Jumbo Privacy [132] and Solid [149]. Despite these approaches covering both the establishment of regulation to limit potential harms (*ex ante*) and techniques to limit the damage after the harms have occurred (*ex post*), both fall upon the individual to fight against the corporations, technology giants, and governments who collect, process, and analyse their personal data. While the implementation of these laws and projects is a step in the right direction, it results in the responsibilisation of data protection from data controllers to data subjects [144], where individuals have the burden of protecting their own personal data as opposed to data controllers themselves. Further, the focus on individual protections and safeguards disregards the power imbalance that lies between users as data subjects and the large corporations as data controllers [65]. Individual data subjects have to exercise their rights against data controllers who are protected by institutional adoption of data protection law and any protest against the data controller's actions requires filling complaints towards the relevant Data

Protection Officer (DPO). Given that individuals and and groups of individuals are impacted by data-related harms, they both should have the option to individually or collectively engage in and collaborate on data protection solutions.

More recently, to address some of these stakeholder tensions and identify the differences in power between them, data stewardship (the process by which individuals or teams within data-holding organisations are empowered to actively initiate, facilitate, and coordinate data towards the public interest) and data governance (a system governing the norms, principles and rules surrounding various types of data) frameworks such as data trusts (applying trust law to data), data cooperatives (managing data through cooperative incorporation), and data collaboratives (applying collaborative frameworks to sharing data) have attempted to provide data subjects with more agency over to what extent their personal data is used [5]. While these frameworks are useful for defining or adapting new legal structures in which data subjects can have additional safeguards over their personal data beyond data protection regulations alone, these solutions may take a long time to implement given that they may not be appropriate to adopt within existing organisations and institutions [217]. Their broad applications and widespread theoretical adoption have also resulted in varied definitions and so require further disambiguation from each other to implement [224]. Although there are current initiatives that aim to standardise and produce practical guidance on how these data stewardship mechanisms could be implemented [48], these mechanisms are not all focused on data protection. They may be focused on data sharing and increasing the value of data through privacy-preserving means without consideration of supporting data subject recourse in cases of data breach or the manifestation of data protection harms [48]. Crucially, these processes may not include data subjects in the iterative process of adopting, building, and deploying the framework to co-create data protection solutions [232], and still result in the responsibilisation of the data protection process.

To address these data-related governance challenges, the commons, a framework that centers around individual and group collective action, trust, and cooperation [174], has been considered to limit the spillovers created by the reuse of data, so increasing its value over time [44]. Central to governing the commons is recognising polycentricity, a complex form of governance with multiple centres of decision-making, each of which operates with some degree of autonomy [178]. The commons can therefore act as a consensus conference [8] to encourage

dialogue among data subjects, experts, policy-makers, and ordinary citizens, creating new knowledge together for the common good. While the adoption of commons principles can help limit the responsibilisation of the data protection process through encouraging stakeholder engagement to co-create data protection solutions, traditionally, such data commons focus on data distribution and sharing rather than data protection [81]. Additionally, data commons with a focus on data protection have only been theoretically applied and categorised through the creation of a set of design principles for data stewardship in a commons [6]. As a result, it is unclear how a data protection-focused data commons can be created in practice and whether it is appropriate for mitigating the responsibilisation of the data protection process through the co-creation of collaborative data protection solutions.

In this thesis, we address the following research questions:

RQ1: Is a data protection-focused data commons appropriate as a socio-technical framework for data stewardship?

RQ2: Can a data protection-focused data commons support the co-creation of data protection solutions for the benefit of data subjects?

1.1 Thesis statement

We make the following thesis statement:

> A data commons created with the aim of protecting personal data can encourage data subjects to co-create and collaborate on data protection solutions, increasing awareness of different stakeholder interests as enabled by data protection law.

In support of this, we make three contributions. We demonstrate that:

1. Existing data commons can be established with a data protection focus, creating a data protection-focused data commons, to overcome current data protection challenges by involving stakeholders with different backgrounds and perspectives.

2. Mapping commons principles to a data protection-focused data commons can not only support data subjects through existing protections offered by data

protection laws and technologies, but also support them in exercising their data protection rights.

3. A data protection-focused data commons encourages more awareness regarding data protection and the use of data subjects' own personal data, allowing them to make choices that are more in line with their own preferences.

1.2 Outline

This thesis is structured as follows:

Chapters 2 and 3 outline the research context, and the state of the art.

- Chapter 2 introduces the digital privacy landscape we currently inhabit, data protection laws we will incorporate as part of the commons, and the data protection challenges associated with existing privacy regulations, technologies, and policies.
- Chapter 3 examines the recent literature that attempts to resolve some of the challenges we identified in Chapter 2, noting the open problems we will address, and introducing the commons.

Chapters 4, 5, and 6 describe our three contributions to support our thesis.

- In Chapter 4, we interview commons experts to identify the data protection challenges for creating a commons, how to overcome them, the feasibility of creating a data protection-focused data commons, and how to improve the commons more generally. The chapter addresses whether a data protection-focused data commons can be created based on existing multidisciplinary experience.
- In Chapter 5, we map out the requirements of a data protection-focused data commons based on our interview findings and existing commons principles, explaining how these could be applied to support data subjects in protecting their personal data. A policy scaffolding is created based on the themes we identified in the previous chapter to establish the data protection-focused data commons as a socio-technical means for data stewardship.

- In Chapter 6, we conduct a study to demonstrate how a data protection-focused data commons can support data subjects to improve their understanding of how their data is used and could be protected through the user study of giving consent to tutorial recordings in context of online learning. In this chapter, we assess whether the data protection-focused commons enables the co-creation of data protection solutions for the benefit of the data subject.

Finally, we conclude with a summary of the contributions we have made, and outline directions for future research.

In summary, this thesis shows that a data protection-focused data commons as a socio-technical framework can support the collaboration and co-creation of data protection solutions for the benefit of the data subject.

Chapter Two

Background

In this chapter, we outline how the legal and technological data protection landscape has changed in context of our data-driven society that we examine in this thesis. We also discuss the challenges that have arisen as a result of these developments. In particular, we explore how existing solutions have led to the responsibilisation of personal data, whereby data subjects (those about whom personal data is collected) have the burden of protecting their own personal data, through case studies.

2.1 Data in our data-driven society

With recent cases such as Clearview AI [116] and Cambridge Analytica [31], personal data is being shared, and occasionally leaked, across different industries to make decisions for and about individuals and groups, sometimes without our knowledge. As part of this data-driven society, different stakeholders have different roles and responsibilities when it comes to our data. Individuals and groups provide or have their data extracted. This is often done by companies and private entities but may also be done by governments and public organisations. Beyond collecting this data, governments in particular have the public responsibility for protecting personal data balanced with public interests such as public safety, public health, and urban planning. Despite data protection's rise to prominence as a result of the increasing number of data breaches and privacy scandals, this concept is not new. The significance of privacy and data protection has only been heightened by technology and by the regulations that govern the innovations, where extensive data collection, management, and sharing practices

may exacerbate power imbalances and tensions between those who build such technologies and its users.

2.2 Data protection and privacy

Beyond the protection of one's own data, privacy is important as it has value for both the individual and for society, protecting us from abuses by those in power, even if we were doing nothing wrong at the time of surveillance [206]. The protection of privacy can be ensured through the law, particularly from data protection law. This is especially true with regards to personal data. Privacy can be broadly described as to encompass "freedom of thought, control over one's body, solitude in one's home, control over personal information, freedom from surveillance, protection of one's reputation, and protection from searches and interrogations" [213]. While something that is private to one person may not be considered private to another, there is still a common understanding of privacy as *protection against* someone or something. The development of privacy as a concept has changed over time, representing how our values as individuals have changed with society. Our perception of privacy has adapted according to our new online environment. Edwards describes these transformations of privacy as being driven by technological advancement, namely the evolution from Warren and Brandeis' [258] right to be let alone (Privacy 1.0) to controlling what personal information is allowed in the public sphere as discussed by Westin [262] (Privacy 2.0) to the current stage of Privacy 3.0, controlling how this personal information is used if it is in the public sphere [65], also modernising Roessler's dimensions of privacy [199]. In the European Union (EU) context, the right to privacy historically dates back to World War II, where the German Constitutional Court in West Germany in 1970 approved what was considered the first modern data protection regulations for public sector data [259]. This section explores how privacy and data protection has become protected in regulation alongside technologies that aim to support the protection of personal data. We then contextualise privacy within our existing data-driven society and explain some of the identified problems that have emerged.

2.2.1 Data protection regulation

Given the historical importance of privacy in the European tradition, laws and regulations have been implemented to protect personal data. Replacing the

Data Protection Directive (DPD) [71], the GDPR [73] came into force on the 25th May 2018. The DPD was introduced in 1995, and with the rise in international processing of big datasets and increased surveillance both by states and private companies, a new regulation was required to modernise and harmonise data protection across EU Member States, irrespective of a data subject's nationality or residence (GDPR Recital 14) [73]. The GDPR, like the DPD, enshrines data protection as a fundamental right under the Charter of Fundamental Rights and Freedoms Article 8 [72]. Although many GDPR rights already existed under the DPD, the new Regulation introduces significant changes as a new data protection framework automatically applied to EU Member States. As with the DPD 1995, the GDPR provides data subjects with rights to exercise against data controllers (those who collect and determine what this data is used for) but does not explicitly provide instructions on how to do so. These data subject rights include the right of access by the data subject (Article 15 RoA, the right to obtain confirmation and access to several categories of information from data controllers about whether the processing of their personal data occurred), the right to be forgotten (Article 17 RtbF, the right to obtain from the controller the erasure of personal data), the right to data portability (Article 20 RtDP and the only new right under the GDPR), and the right not to be subject to a decision based solely on automated processing (Article 22). Data controllers may have DPOs who deal with GDPR requests.

The GDPR, like the DPD, provides several rights with regards to personal data for data subjects as "identified or identifiable natural" (Article 4(1)) persons to exercise against data controllers as those who determine "the purposes and means of the processing of personal data" (Article 4(7)). Unlike the DPD, data controllers' responses to data subject rights requests shall be provided free of charge (Article 12(5)) unless manifestly unfounded or excessive (Article 12(2)). Additional information about the data subjects may be asked to enable their identification if there is reasonable doubt about their identity (Article 12(6)). When further information and proof of identity is received, data controllers cannot refuse to act upon the data subject's request (Article 12(2)). Once confirmed, the data controller has up to one month, or up to three months if the complexity and number of requests are significant, and without undue delay to comply (Article 12(3)). While the GDPR was implemented in recognition of rapid technological developments, the Regulation aims to be technologically neutral and not depend on the techniques for the protection of natural persons (Recital 15). Instead, the GDPR has introduced qualified duties to principles such as Data Protection by

Design (DPbD) [251] and avenues to certification for regulation innovation in data protection markets [98]. Further guidance is offered by each Member State's Data Protection Authority (DPA).

In addition to the GDPR, other regulations such as the United States (US) CCPA [32] attempts to rebalance power between data subjects and data controllers. Information rights such as the right to access to recorded information held by public sector organisations through the Freedom of Information Act [245] and intellectual property rights offered under the Directive on Copyright in the Digital Single Market [74] can further support data protection for data subjects in offering greater transparency and alternative remedies to issues relating to personal data. Outside of the US and the United Kingdom (UK), other jurisdictions also protect privacy and the right to privacy in different ways. Koops et al. analysed the constitutional protection of privacy in nine primary countries: the US, UK, Canada, the Netherlands, Germany, Italy, Czech Republic, Poland, and Slovenia [134]. The authors also touch upon other jurisdictions such as Argentina, Brazil, Chile, Russia, South Africa, and Uruguay. While countries such as Slovenia, the Netherlands, and Poland directly reference privacy and right to private life, countries such as Germany, Italy, and Russia protect privacy through a combination of constitutional rights, personality rights, and the protection of the home and mediated communications. Argentina and Uruguay protect private actions as opposed to private life. Estonia, Japan, and South Africa provide privacy protections in relation to property of persons.

Although the UK has transposed the GDPR to its own data protection law under the Data Protection Act [244] as governed by its own authority, the Information Commissioner's Office (ICO), this thesis focuses on the GDPR due to its extraterritorial applicability across Europe and around the world. As of the date of writing, there are only minor differences between the GDPR and the Data Protection Act, such as safeguards for journalists. There were also proposals by the UK Government to scrap certain GDPR protections after the UK's withdrawal from the EU (Brexit) on 31 January 2020 "to make the country's data regime even more ambitious, pro-growth, and innovation-friendly" [59].

It is useful to note that while data protection and privacy are used interchangeably in this thesis, data protection focuses more narrowly on the protection of personal data, where privacy addresses theories surrounding our rights to private life. Notably, the word "privacy" is not used at all in the GDPR. This thesis also does

not refer to data ownership as part of data protection or privacy. Narrowing down the scope of the GDPR, we will focus on applying data subject rights, with references to wider data protection principles such as DPbD, throughout this thesis.

Overall, the GDPR has been viewed as a modernisation of the previous data protection regulation applied to our data-driven society. The new Regulation aims to prevent and circumvent the abuse of personal data across politics, industry, and government, with overarching legislation on how systems that require personal data should be developed, assessed, and governed both prior to its deployment and continued use.

2.2.2 Data protection and technology

In recent years, the law has attempted to modernise its regulation of data and data protection to keep up with the rapid technological developments that gather data. Although technology has facilitated the widespread collection of personal data, a growing body of technologies have focused on providing more transparency and options for individuals to see how their personal data is or is not used.

Technology-driven tools have been developed to help data subjects and data controllers exercise and comply with the GDPR. Some tools include, DataStreams.io [51] (a consent manager for data controllers and a data stream manager for data processors), the Data Transfer Project [231] (an open-source, service-to-service platform that facilitates direct portability of user data between cloud services by converting proprietary application programming interfaces (APIs) to and from a small set of standardised data formats, founded by Facebook, Google, Microsoft, and Twitter), Fair&Smart [76] (an application that helps French data subjects claim GDPR rights, regain control of their privacy, and decide who to trust with their personal data, while supporting GDPR compliance and management services for data controllers), My Data Done Right [23] (a project that helps Dutch data subjects exercise the RoA and RtDP), DoNotPay [62] (a legal services chatbot that sought claims from Equifax for its security breach [17] as one of its services), and Jumbo Privacy [132] (an application that allows data subjects to backup and remove their data from platforms, and access that data locally). New technologies have also attempted to give users the ability to control their own data more broadly. Some tools include Databox [45] (a personal data management platform that collates, curates, and mediates access to an individual's personal data by

verified and audited third-party applications and services) and Solid [149] (a decentralised peer-to-peer network of personal online data stores that allows users to have access control and agency over the storage location of their own data). These tools can be useful as they reveal and create new forms of granular control for data subjects over their own personal data while supporting centralised forms of data management for data controllers through the tool's dedicated platform.

Some technological tools aim to help data controllers and data processors, such as OpenGDPR [173] (an open-source common framework that has a machine-readable specification, allowing data controllers and data processors to communicate and manage data subject requests in a uniform, scalable, and secure manner) and Port.im [187] (an application that provides data controllers with the possibility of connecting applications together, creating GDPR compliant agreements, communicating how personal data is stored to data subjects, and taking control of the data stored). De Hert et al. also consider the possibilities for building interoperable infrastructures enabling data subjects to bridge the gap between specific services [52]. These tools enable data controllers to more easily standardise their data protection practices, including automating their data protection-related interactions with data subjects.

When created following GDPR requirements, these tools can guide what rights data subjects have without them having to have read the law. As the tools are developed by third parties, data subjects do not need to self-manage the process. For data controllers, tools do not absolve them of their responsibility to inform data subjects about their rights but can lower the resources dedicated to replying to requests. There may also be greater certainty and consistency with the standardisation of responses. However, despite the existence of these regulatory technology tools, many organisations continue to take a manual and informal approach to GDPR compliance [202].

In this section, we introduced the data protection laws and technologies that aim to protect data subjects' personal data. The next section considers the data protection challenges that may be insufficiently addressed by these legal and technological instruments.

2.3 Data protection challenges

While the GDPR is a step forward for the protection of personal data, it is insufficient on its own because, by nature with its traditional and legislative approach, it takes a long time to come into fruition. The language of the law and legalese can be convoluted and difficult to understand. This may also be true of data protection technologies, where only those familiar and are comfortable testing new digital tools will be able to benefit from them. In this section, we outline some of the interdisciplinary challenges of protecting personal data in context of the current data protection landscape.

2.3.1 Regulatory challenges

As outlined in Chapter 2.2, the protection of personal data in our digital environment has been facilitated through legal and technical means. However, these solutions are imperfect, particularly due to the lack of power data subjects have compared to the increasingly sizeable and international data controllers that collect, analyse, and share their personal data.

Before considering these stakeholder tensions, one of the challenges to protecting personal data is the technological neutrality of the GDPR. To preserve the longevity of the GDPR and limit circumvention, the Regulation was created with the intention of being technologically neutral (Recital 15). However, this may hinder the protection of digital personal data given the technical requirements needed to ensure protection [114]. More specifically, the GDPR is insufficient because it is *ex post* in its protection of personal data, only predominantly providing redress to data subjects after such data has been put forward. For example, in considering the RtDP for data subjects exercising their rights, it is unclear whether there is too much onus on them to right any wrongs when there is discord between different GDPR rights. Referring to the then-draft GDPR, Swire and Lagos argue that data portability may reduce consumer welfare as it places excessive burden on small and medium enterprises by disregarding market power and efficiencies [226]. Security challenges may arise as the complexity of controlling and processing personal data increases with more portable data [260]. Further, Graef et al. argue that if justifications for data portability are poorly-defined, portability may be considered a goal in itself for data controllers, with little impact on personal data protection [97]. Additionally, data subject rights and DPbD may conflict when exercised and deployed [251]. This is because as a GDPR

requirement, DPbD may be restrictive in practice, emphasising privacy-as-control over privacy-as-confidentiality when no data protection frameworks guide DPbD employment. The privacy-protective RtbF may also override the RtDP as a result of "multiple linking", where two or more data subjects can be easily linked by same datasets [138]. For data controllers, the GDPR rights offered to data subjects may be considered overly onerous to respond to. For example, the lack of clarity on authenticating the data subject's identify could result in potential data breaches or denial of access [25]. The content of the rights themselves may include information dating back to before the legislation, retroactively requiring data controllers to respond. The exercise of rights by data subjects may also amount to the abuse of the GDPR. Data subjects may request for their data against the data controller due to disagreements with actions unrelated to the regulation, such as the Activision Blizzard's player ban [246], who must respond even though the data subjects do not actually want access to their data.

Crucially, the existence of data protection laws does not mean that people are able to exercise their enshrined rights. Although individuals are more aware of their data subject rights now compared to pre-GDPR, these are not well understood [165]. Only 15% of EU citizens indicate that they feel completely in control of their personal data [46]. Evaluating location-based services, Herrmann et al. found that individuals do not necessarily know all the inferences that are made using their data and thus do not know how it is used [111]. Importantly, individuals are unaware of and unable to correct false inferences, making the collection, transfer, and processing of their location data entirely opaque. With privacy policies written in legalese and privacy-protecting options hidden beneath dark patterns (user interface features that are designed to intentionally manipulate, steer, or nudge users into actions that may not be in their best interest) [99], data subjects cannot easily find out how their data is reused, aggregated, and anonymised to make decisions about them [248]. As a result, data protection solutions should not be wholly reliant on legal infrastructures and mechanisms, where legality alone may not provide data subjects with the confidence that their personal data is adequately protected.

2.3.2 Stakeholder tensions

Even if enforcement were stronger, another reason for the difficulties in studying the relationship between data protection, technology, and the law is because of the

large numbers of stakeholders involved. Tensions arise between the data subject, the technologist developing new technologies, the enforcers, and policy-makers, where different stakeholders have different motivations and interests. Table 2.1 represents a taxonomy of stakeholders that may be involved within data privacy and protection cases as defined following the GDPR, including a description of their roles. The table demonstrates how different data protection stakeholders regard their relationship to personal data and how their different motivations can bring forward tensions that impact how personal data and other stakeholders are treated.

Data Protection Stakeholders	*Description*	*Motivation*
Data Subjects	Data subjects (individuals or groups) as defined under the GDPR.	Seek the right balance between sharing data for generating personal and common value with their own personal privacy preferences.
Data Controllers	Data controllers as defined under the GDPR.	Maximising the use of data subjects' data for generating economic value while adhering to regulations and data subject preferences.
Data Processors	Data processors as defined under the GDPR.	Process data according to the needs of data controllers.
Large, Technology-driven Companies	Big technology companies such as Google, Apple, Facebook, and Amazon that have larger capacities for dealing with data, often as both data controller and processor.	As data controllers, maximising the use of data subject data for generating economic value while adhering to and lobbying for regulations that suit company aims.
Data Protection Authorities (Enforcement)	DPAs as defined under the GDPR.	Enforce the GDPR while balancing the tensions between different stakeholders within the local and global political environment.
Data Protection Authorities (Norms)	DPAs as defined under the GDPR.	Pushing for further guidance and development within and beyond the data protection landscape.
Law and Relevant Supervisory Bodies	Groups that specifically pertain to the law or have a role in supervising organisations that fall within their remit.	Provide independent guidance and insight into how the law is to be navigated.
Standards Bodies	Groups such as the European Data Protection Board that have the power to set standards of practice but do not have full legal force.	Balance political appetite with business and regulatory changes to put pressure on data protection authorities and lawmakers to enforce the law in certain ways.
Civil Society	Organisations such as Open Rights Group, Privacy International , Art19, Electronic Frontier Foundation, and RightsInfo that focus on furthering social causes outside of government and businesses.	Challenge the existing norms surrounding the law in pursuit of organisational agendas, often going against big technologies companies.
Policy Makers	People tasked with developing and writing policy that could be in government, part of enforcers or authorities, or be independent.	Balance the needs of voters, national agendas, and international economic outlook.

Table 2.1: Data protection stakeholders and their general motivations in this area. The variety of stakeholders and their contrasting motivations illustrate potential tensions related to the governance of data.

The broad range of stakeholders illustrates the complexity balancing data protection priorities and motivations to minimise tensions. Many stakeholders play dual roles in terms of the power relations and tensions between them. For example, data controllers can act in both positive and negative ways, developing tensions with data subjects by nature of their relationship with the personal data between them. While enforcement-focused DPAs may have the data subjects' interests at heart, data subjects may end up giving away more of their personal data to them during investigations for purposes such as identity verification, resulting in fewer privacy and data protection measures. As a result, additional forms of data management and stewardship, a concept that embodies the responsible planning and management of common resources [7], may be considered to balance stakeholder tensions to ensure that these institutionalised challenges from different data protection roles are collectively acknowledged and solved.

2.3.3 Enforcement

Enforcement of the GDPR has also proven to be difficult. Laws focusing on placing data protection responsibilities on data controllers and empowering enforcement bodies assume that data controllers understand how to implement those responsibilities and that enforcement is successful. DPOs' and DPAs' enforcement practices are inconsistent and unclear due to lack of guidance [165]. Data controllers responding to GDPR Article 20 RtDP requests provided a large variation of file formats that were not all GDPR compliant and confused the right with other data subject rights [268]. Kamarinou et al. found inconsistencies in details and lack of transparency about third-party storage and the processing of personal data of cloud service providers' in their terms and privacy policies [133]. There have also been enforcement issues related to GDPR's One-Stop Shop mechanism (Recital 127), which allows data controllers that engage with EU cross-border processing to only have to deal with a single lead supervisory authority for most of their processing activities [127]. The UK was previously also part of this arrangement until Brexit [124]. The lead supervisory authority is represented by the DPA where the company's main establishment is registered (Article 56(1)). For many large, multinational companies such as Facebook and Google, this is often the Irish Data Protection Commission given Ireland's low corporate tax rates [123]. Due to this political tension between wanting these companies to remain in Ireland and the enforcement of data protection regulations, the Irish Data Protection Commission has been lenient in their decisions. For example,

Facebook was fined US$36 million on one occasion, a fine that would take the company just over two and a half hours to earn in revenue [140]. Concerns over Ireland's application of the GDPR have been recognised by the European Commission's president and EU ombudsman following the Irish Council for Civil Liberties' complaint over the country's lack of enforcement over big tech investigations [228]. Further enforcement challenges are compounded when funding for DPAs may also be limited, especially in comparison to the large corporate data controllers. This is true particularly for the case of Ireland, where the Irish Data Protection Commission was given only 27% of its requested increase by the Irish Government, totalling €21.1 million, despite increased responsibilities post-GDPR [227]. Two years after the implementation of the GDPR, the European Commission praised the Regulation for providing national DPAs with the right tools to enforce the rules [70]. However, they did admit that there is still a "very serious to-do list" and that the EU still requires more efficient, harmonised, and vigorous enforcement of the GDPR.

2.3.4 Responsibilisation of data protection

While the tools highlighted in Chapter 2.2.2 are useful if they offer controls that limit the processing of personal data according to data subject preferences, it crucially results in the responsibilisation of data protection from data controllers to data subjects [144], where individuals have the burden of protecting their own personal data as opposed to data controllers themselves. In the case of data infrastructures, these tools may not be able to solve challenges related to the management of data, and it is unclear how they would meet GDPR requirements [24]. Existing tools also frame privacy as control by placing individual onus on data protection, without supporting other GDPR principles such as DPbD or data minimisation. Tools may assume that data subjects already have a sufficient level of understanding of their data subject rights by focusing on more fine-tuning privacy settings and features. They also require data subjects to trust the companies and the technological services they provide [61]. While there is awareness from data subjects of how their data protection rights could protect their personal data, there are frustrations on the perceived regulatory burden related to the regulation of day-to-day behaviour of individual citizens as a result of how the GDPR is operated by legislators, data protection authorities, and data controllers in practice [223].

From the data subject's perspective, there may also be downsides to using these tools for exercising their GDPR rights. As none of these tools are certified by GDPR governing bodies (Article 42), there is no guarantee that using them will help data subjects exercise their rights or ensure that data controllers are compliant. Some tools charge fees to data subjects and data controllers for what should be a free right. For example, although not directly mentioned by the company, FreeYourMusic (an application that ports music playlists from one platform to another) effectively acts upon your RtDP right to port your data from one data controller to another by charging a fee [85]. Although tools may simplify the process, there are many tools on the market, conversely overcomplicating the process for both data subjects and data controllers. Tools may confuse data subjects with no single way of exercising GDPR rights. While third-party tools exist, additional privacy and security issues could emerge from using these services through the transmission and processing of additional personal data. The data contained in data subject access request responses may be inappropriate for sharing with third parties. The private companies such as advertising technologies (AdTech) companies operating the tools may monetise personal data passed onto them when used as an intermediary for exercising GDPR rights. As a result, even though some tools were created to support data subjects with exercising their rights, data subjects still have to do their due diligence to check whether those tools comply with the GDPR and how their personal data is being processed by these third parties.

Additionally, despite data breaches and privacy scandals affecting large groups of individuals, there are no tools that focus on supporting data subjects in groups *ex ante* when it comes to finding collective data protection solutions. Even when there are *ex post* group and collective solutions, such as class action cases or unionisation, significant power imbalance between individuals and large, multinational companies remain. Given that for data controllers, personal data is aggregated and used to generate economic value [212], data subjects should have the ability to co-create data protection solutions with other data subjects that moves the focus away from individuals and towards groups, "from processes of consumption to those of citizenship and accountability" [229]. Data protection solutions could consider supporting collaborative data protection, where information gathered from individuals could be shared amongst each other to limit the responsibilisation of the data protection process.

In this section, we considered the data protection challenges manifested from legal and technological applications of data protection requirements. These include stakeholder tensions, enforcement difficulties, and wider regulatory challenges. Even with legal and technological instruments that aim to simplify the data protection process for data subjects, they may still be burdened by having to protect their own personal data given the focus on individual rights and protections with respect to data protection. In the next section, we examine case studies that exemplify the responsibilisation of the data protection process, and why engaging with collaborative solutions may decrease the individual burden of data subjects.

2.4 Case studies

To demonstrate the impact of the responsibilisation of the data protection process and other data protection-related challenges, this section explores three case studies: Cambridge Analytica, the use of data in England's National Health Service (NHS), and education data and online learning.

2.4.1 Case study 1: Cambridge Analytica

The Facebook and Cambridge Analytica case illustrates how little individual data subjects can do to protect their personal data, particularly where there are limited viable alternatives to a service. In March 2018, The Observer revealed that 50 million Facebook profiles were harvested for the data analytics firm Cambridge Analytica and were used to build models with the aim of influencing elections [31]. The data came from authorised access through the "thisisyourdigitallife" quiz application developed by Aleksandr Kogan from the University of Cambridge [139]. Kogan was granted permission to access Facebook users' accounts and their friends' information, passing this information on to Cambridge Analytica without user consent. Although only 270,000 people downloaded the app, Facebook believe that up to 87 million profiles may have been improperly shared [208]. Responding to the scandal, Facebook CEO Mark Zuckerberg wrote to the over 2 billion users in a Facebook post that the company has "a responsibility to protect your data, and if we can't then we don't deserve to serve you" [274]. Cambridge Analytica and its parent company Strategic Communications Laboratories were suspended after the news went public [101]. Since the story broke, Facebook removed developer access to personal data if apps were left unused for three months [208], made its privacy tools easier to find [66], and shut down Partner

Categories for third-party advertising [162]. Although only 10 people downloaded the app in New Zealand, the head of communications for Facebook Australia and New Zealand estimated that a total of 63,724 people were impacted [201]. This case demonstrates how our ability to protect our personal data relies on privacy dependencies, the many ways that our privacy depends on the decisions and disclosures of other people [14]. Ultimately, the company was fined the maximum £500,000 possible under pre-GDPR regulations for the data breach [110]. Whilst Zuckerberg has admitted to and apologised for some of these mistakes, even if in earnest, self-regulation, as seen from Facebook's previous privacy blunders, is insufficient. As its business model depends upon sharing personal data for free with developers in exchange for value derived through greater platform power, the company has no incentive to change its existing privacy policy. Cambridge Analytica is significant because of the combination of the re-purposing of data without consent, the inclusion of third parties, and the implications on wider democratic society.

As the Cambridge Analytica case study shows, individuals as data subjects are powerless compared to companies as data controllers that are able to create vast data networks and monetise that data as part of their business model. Importantly, even if these companies breach data protection regulations, the resulting fine is so small even at the maximum 4% of annual turnover rate of the previous year as required under the GDPR (Article 83). The economic gain from gathering significant amounts of personal data outweigh the cost of the fine, with no incentive for these companies to protect personal data. As a consequence, irrespective of whether data subjects can exercise their rights or use tools to support their data protection preferences, the power imbalance between them and the data controller is not surmountable as an individual.

2.4.2 Case study 2: National Health Service

Beyond the private sector, data subjects may also have little input and have difficulty challenging data controllers' personal data practices as individuals in the public sector. Public services have also been datified in ways that are not always clear to the public. Using the health sector as an example, launched in 2013, NHS England revealed its plans for the care.data programme. The programme aims to extract individuals' medical records from General Practitioner (GP) doctor practices, linking information from various NHS providers to obtain "a more

complete picture of how safe local services are, and how well they treat and care for patients across community, GP and hospital settings" [67]. However, the project was temporarily suspended for six months in February 2014 after concerns were raised when NHS England was unable to explain the benefits of care.data [93]. Importantly, it was unclear how care.data would handle and protect sensitive personal data. It was also unclear how this aggregated data could directly benefit local communities. A public campaign aimed to educate the public and provide them with the option to opt-out was deemed inadequate in its delivery and impact by the British Medical Association (BMA) and the Royal College of General Practitioners (RCGP) [92]. With a brief start in June 2015, the programme was again suspended after the Secretary of State for Health at the time requested the National Data Guardian (NDG) for Health and Care, Fiona Caldicott, to review data security, consent, and opt-outs for the NHS and social care which was published in July 2016 [104]. The Care Quality Commission (CQC) also issued a "Safe data, safe care" report on whether personal health and care information was being used safely and appropriately protected in the NHS [39]. Immediately after their publication, the care.data programme was withdrawn permanently as the report suggested stronger consent and opt-out measures than NHS England initially envisaged [27]. However, more recently, in May 2021, NHS Digital proposed the launch of the General Practice Data for Planning and Research (GPDPR) scheme, where England's GP health data, with identities partially removed, would be made available to researchers and companies for healthcare research and planning [128]. The deadline was initially delayed to September 2021 but the GPDPR is now currently on hold with no date for implementation, following doctors' and privacy campaigners' push-back, with over 1.3 million individuals opting out by June 2021. Despite initial push back on the care.data scheme, the redeployment of a similar programme in the form of the GPDPR indicates the lack of power individuals have when it comes to limiting the collection, use, and re-purposing of their own personal data.

As the care.data and GPDPR case study demonstrates, it was group and collective action that resulted in the suspension of those programmes. Even though everyone can be considered data subjects given that the NHS is the public health system in England and the UK, without formal collective action from influential groups, it would have been difficult for individuals to not only understand the implications of aggregating personal health data but also whether and how they could opt-out of the programmes. Notably, it was only with input from the BMA, RCGP, NDG,

and CQC that the care.data programme was suspended due to the organisations' concerns regarding personal data, where public interest was also of concern beyond data protection alone. As similar organisations and governing bodies are not required to present their views when it comes to assessing data protection practices, data subjects may miss out on important information from external and independent experts that can help inform their decisions about the potential data-related harms individuals could face. As a result, collective pooling of information and data as well as having access to experts in the field could benefit all stakeholders in understanding what data protection concerns may exist.

2.4.3 Case study 3: Education data and online learning

The final case study we explore to demonstrate the responsibilisation of personal data is education data and online learning in higher education. The integration of technology into education, or education technologies (EdTech), has long existed. From e-mails to using laptops in the classroom, technology has allowed for more flexible and inclusive ways of learning while introducing new methods for collaboration and information sharing [95]. However, technological developments have also increased the responsibilities that institutions have over student data, expanding and blurring the lines of what education data entails. Education data has been described by Borgman as "grey data", where the lines of research, teaching, learning, services, and administration data blur in the context of universities [26]. This raises questions around privacy and data governance, where academic data collection has become more complex due to the open access requirements of institutions, as well as the accumulation of such grey data in daily academic life. The governance of online learning data is further complicated as private and corporate stakeholders use this data for third-party purposes, such as assessing university rankings, exam proctoring, or graduate surveys.

The digitisation of education has resulted in greater data collection, storage, and analysis through learning analytics. While learning analytics can help institutions understand student engagement, improve teaching, and improve the overall student experience [131], they have similar characteristics to big data and so have similar data protection concerns, particularly regarding relationships between universities as data controllers and students as data subjects [41]. In particular, during the COVID-19 pandemic, the digitisation of education increased exponentially as institutions moved all their teaching, research, and administration

services online. While higher education institutions (HEIs) have done their best to ensure that online learning is conducted in a safe and secure manner, the digitisation of education has resulted in more data-related harms. From "Zoom-bombing" (where a person joins a Zoom meeting uninvited and aims to disrupt the session) [18] to monitoring [221], students have been negatively impacted by these new technologies, resulting in potential harms that impact their lives beyond academia. Reflecting on online learning experiences, in a study by the Center for Democracy & Technology, 56% of primary and secondary students were found to be concerned about privacy and security of student data and information collected by their school [34]. 82% thought it would be helpful to know which privacy controls to set and how to do it, and 79% thought it would be helpful when choosing an app to be able to tell which ones do a bad job of protecting personal data [34]. The report also found that the more senior the student, the more likely they were to have concerns, demonstrating how these considerations may also be taken in context of higher education.

E-proctoring, or the use of virtual proctoring software to monitor students through webcams, microphones, and other tracking tools with the aim of preventing cheating, has also become more commonplace. The use of e-proctoring technologies could harm agency and trust [77], as the surveillance environment created is counter-productive to learning [243]. Other concerns include the added stress of being monitored [82], the software being incompatible with devices [107], and the time taken to implement [119]. It is also unclear whether proctoring can achieve its purpose in preventing cheating [15]. In one example, a student exercised their GDPR Article 15 RoA to see what data the proctoring software was gathering about them. They found that many incidences flagged as "audio level in the room was above threshold" and "the test taker looked away from the exam page" were full of false positives, especially when staff turned up the sensitivity settings [89]. Algorithmic test proctoring may also discriminate based on gender and race [225]. The use of proctoring services was challenged by UK bar professional training course students, where students were monitored using webcams throughout the examination without any breaks and moving away from the webcam would result in automatic termination [200].

The data protection considerations of tools and the usefulness of lecture and tutorial recordings have also been questioned. Many tools used by HEIs to deliver online learning (such as Zoom and Microsoft Teams) were not created

for education. As a result, these third-party companies may be less sensitive to stakeholders' motivations, where students are treated as consumers, without regard to their participation in education [64]. For example, the Microsoft Office Productivity Score included in Microsoft Teams tracks the time and activity of its users, producing data on the extent to which individuals are working on its platform. Initially, this data could be accessed by institutions and linked to specific usernames. Even if HEIs do not access this data, it could still be collected by digital platforms and may be shared and sold to third parties. Only after privacy concerns were raised did Microsoft remove usernames and change how the data gathered is presented [214]. Particularly where universities are public institutions, these data processing practices should be made transparent to those who use these technologies. Additionally, the reuse of recordings may not be clarified to students. A student at a US university only found out that the professor delivering their online class had died two years earlier when the student tried to email them during the pandemic [1]. Taken to the extreme, there may also potentially be political harm for individuals if the risks of online learning data and recordings are not properly managed, with institutions choosing not to record tutorials discussing sensitive political topics [118] or self-censor to minimise their digital footprint [270].

As the digitisation of education demonstrates, there are numerous data protection risks that come with online learning. Importantly, in response to potential data-related harms caused by education data, collective and collaborative solutions have been studied and created. Many organisations have looked at the impact of the pandemic on digital education. The Office for Students (OfS) engaged stakeholders to produce guidance establishing the essential components of successful digital teaching and learning, recommending core practices HEIs can use to improve online learning for students [167]. The Joint Information Systems Committee (JISC) has written a report to understand the COVID-19 response and explore the future of digital learning and teaching [130]. Policy solutions were also devised for identifying the future role of emerging technologies in education and training [256]. In supporting more inclusive and equitable online learning practices, researchers and practitioners have shared their experiences of online learning during the pandemic [265]. The shift to online learning introduces new questions around the ethics of care related to online and remote work [84]. The Centre for Research in Digital Education created the Manifesto for Online Learning to illustrate how surveillance culture can be resisted [16]. Silverman et al. share their lessons on helping staff transition to authentic assessments without e-proctoring [211].

More broadly, collaboration with students and limiting the impact of responsibilisation can also support agency and trust both in the data protection process as well as with their institutions. Plunkett et al. found that to ensure that student privacy frameworks align with students' digital practices and privacy expectations, adult stakeholders may consider incorporating robust ways for youth to participate in discussions about tackling student data privacy challenges [186]. Teachers have mentioned the importance of students voicing concerns about the use of novel technologies in education [36]. Addressing how this can be done, JISC suggests that universities prioritise blended learning approaches where possible, and that students co-design curricula [79]. Williamson and Hogan recommend that higher education stakeholders should work collegially to define alternative imaginaries that can guide post-pandemic recovery of HEIs, moving away from using academia as an engine for producing measurable learning performance and associated workforce productivity gains [265]. Prinsloo and Slade further create a framework to support learner agency [193], recognising that it is impossible for individuals to comprehend the scope of data that might be collected, analysed, and used, and its implications when it comes to learning analytics [194]. This framework includes contextual integrity [164] of privacy and data (where adequate protection for privacy is considered alongside norms of specific contexts), student agency and privacy self-management, rethinking consent, and employing nudges. Co-created solutions to navigate privacy and security during online learning were also crowd-sourced such as the "Coronavirus Tech Handbook" [161] and "A Comprehensive Guide To Tech Ethics and Zoom Class" [91]. While these collaborative efforts may not specifically address data protection concerns, they all support data management and stewardship approaches to education data. As a result, in order for education-related personal data to be adequately protected and to foster an engaging digital education environment, all stakeholders need to be involved with the co-creation of socio-technical guidelines for online learning, with a clear understanding of how the collected data can be used to benefit students, staff, and institutions.

2.4.4 Case studies summary

The Cambridge Analytica, NHS, and online learning case studies explored in this section demonstrate how the responsibilisation of the data protection process for data subjects makes it difficult to challenge the data controllers and institutionalised data protection practices that may not be beneficial for individuals.

Even if data subjects are able to exercise their rights under the law and set their data protection preferences using technological tools, they may still have expended a large amount of personal time and energy individually to understand their rights or how to use certain platforms. These efforts could be decreased with the pooling and sharing of information, as demonstrated by the crowd-sourced efforts to tackle privacy and security challenges for online learning. Beyond legal protections, community and stakeholder values need to be considered when determining the appropriate data protection solution due to moral and ethical aspects of care between data subjects' and data controllers' roles outside of data protection, such as doctor-patient and university-student relations. This analysis confirms our motivating suspicions that socio-technical requirements as part of wider data management and stewardship considerations could be implemented to encourage the co-creation of data protection solutions for common benefit.

2.5 Summary

In this chapter, we have introduced the privacy and data protection issues that we will consider for the remainder of this thesis, and discussed the interdisciplinary challenges associated with the responsibilisation of data protection and power imbalance between data subjects and other stakeholders. We note the following:

- In our data-driven society, people's personal data is being increasingly collected, processed, and analysed by companies. This can result in privacy and data protection challenges for individuals as data subjects.
- To mitigate some of these privacy issues, data protection laws and privacy tools have been implemented to provide rights to data subjects and protections for their personal data.
- Despite data protection safeguards, data subjects still face significant power imbalances against data controllers as data protection solutions often place responsibility on individuals to protect their own personal data and manage any data protection risks.
- Our case studies demonstrate how the responsibilisation of the data protection process can be limited through collaboration and co-creating solutions between stakeholders, suggesting alternative socio-technical forms of data management and engagement to support data protection preferences.

In the next chapter, we consider recent research that aims to tackle the privacy and data protection issues we identified when it comes to protecting personal data, and examine proposed data management and data stewardship solutions to the responsibilisation of the data protection process.

CHAPTER THREE

CURRENT STATE OF THE ART

In this chapter, we explore recent multidisciplinary research that attempts to resolve the data protection challenges as discussed in Chapter 2 and identify collaborative solutions. We also survey data stewardship literature to determine the extent to which work in this domain tackles the issue of the responsibilisation of the data protection process. We outline how many existing data stewardship frameworks focus on "data", with little direct and ongoing data subject engagement and participation considered as part of the stewardship process. As a result, we assess how a commons can be applied to data protection to facilitate collaborative pooling of resources and co-create data protection solutions centred around data subjects.

3.1 Resolving data protection challenges

In Chapter 2.3, we outlined the data protection challenges identified in the literature with reference to data protection laws and technologies. In this section, we discuss the broader context for these issues and the emergence of interdisciplinary solutions.

We noted in Chapter 2.3.4 that the responsibilisation of personal data is not directly addressed through current laws and technologies. The protection of data needs to go beyond these data protection frameworks and tools, focusing on collective and collaborative solutions to overcome the burden of individual responsibilities within the data protection process, as identified more generally in Chapter 2.2.

Privacy is a multi-faceted concept and encourages technical, legislative, and socio-political techniques that can be used by data subjects, data controllers, and data processors. Lessig, in his seminal work *Code: And Other Laws of Cyberspace, Version 2.0*, wrote that the regulatory modalities of law, architecture, markets, and norms are all necessary when considering the governance of cyberspace [136]. Architecture, in the context of our online environment and defined as technological constraints, represents the code which data science projects are built upon. Lessig argued that "code is law". The self-regulatory nature of code is *ex ante* and differs from that of laws, markets, and norms because the latter are subject to a discourse with those subject to their control. However, Reed and Murray argue that Lessig's model is a misunderstanding of the ways regulatory modalities interact [198]. Lessig's model, while accurate in its identification of modalities, does not go further to represent the complex network of interactions between individual cyberspace users, code producers, lawmakers, and markets. Additionally, Lessig's model also fails to account for conflicts that relate to the wider community and their reaction towards regulatory modalities they disagree with. Given the multidisciplinary nature of data protection as outlined in Chapter 2, the protection of personal data could acknowledge these regulatory modalities and their complex relationships. Considering the regulatory dimension more specifically, the GDPR should not be considered as a one-size-fits-all piece of legislation for governing technology and the data associated with it. Interaction with other regulatory levers and bodies beyond data protection may support a more holistic understanding of regulating such harms [109].

More practical applications of data protection laws and principles can also be supported beyond regulation. For example, applying concepts such as contextual integrity [164] can help demand that information gathering and dissemination be appropriate to that context and obey the governing norms of distribution within it. This also shifts the responsibility for considering data protection issues away from the data subject and towards critical applications of community and common practices within use cases, as identified from the case studies in Chapter 2.4. Other disciplines such as political sciences and Science and Technology Studies (STS) have also addressed challenges to responsibilisation through the "problem of many hands" [238] and "cooperative responsibility" [108] respectively. Using economic and environmental examples, Thompson illustrates how the problem of many hands requires designated overseers to be held responsible for monitoring organisational structures and to make the necessary, iterative changes to strengthen

individual and community responsibility for the future [239]. This is due to the difficulty of assigning responsibility in organisations where many different individuals contribute to decisions and policies. Given that it may be difficult to identify who is responsible for collective harm in particular, both backward- and forward-looking responsibilities should be identified, where a more specific and restricted problem is addressed to achieve a desirable collective outcome [249]. The greater moral responsibility may then be attributed to those with greater power through cooperative responsibility, where the realisation of public values in platform-based public activities could be the result of dynamic interaction between platforms, users, and public institutions [108]. This is expressed in forms such as organisational and design responsibility for platforms, active participation, empowerment and real responsibility for users, and creating frameworks for shared responsibility and shared values for governments. As a result, platforms and users are considered as partners in regulation rather than as subjects. While it is acknowledged that cooperative responsibility may be in conflict with additional duties of care for platforms, the necessity for user empowerment may be considered given the imbalance of power between them and other stakeholders.

Further, careful considerations of agency and reflexivity through the notion of voice, a value for social organisation that takes into account agents' practices of giving an account of themselves and their conditions of life, can help foreground notions of accountability in data calculation, ownership, and use [42]. This is particularly important as transparency, explainability, and accountability within governance alone do not adequately address power asymmetries and the underlying architectures that support such data gathering [188]. To resolve the challenges resulting from the responsibilisation of personal data identified in Chapter 2.3.4, we could look beyond the GDPR's safeguards. Solutions and frameworks that address how data is managed and governed within existing data infrastructures can be identified, supporting the protection of data through greater transparency, accountability, and collective responsibility.

3.1.1 Human-centred design in technology

As we introduced in Chapter 2.2, as individuals become more aware of privacy issues, other methods of addressing data-related harms such as policies, human-centred design, and privacy-enhancing technologies (PETs) have been applied.

Research and tools have implemented human-centred design to improve the relationship people have with technology, particularly in relation to limiting the stakeholder tensions identified in Chapter 2.3.2.

Focusing on data and technological governance, codes of ethics and conduct have been created and updated to reflect changes in our data-driven society. For example, the Association for Computer Machinery (ACM, the world's largest educational and scientific computing society) Code of Ethics was updated in 2018 to include algorithmic transparency, informed consent, addressing discrimination, and respecting privacy [96]. Research has also been done to identify and mitigate the adoption of dark patterns [99]. This includes embedding ethics into the user experience design process [37], critiquing the implementation of consent banners [100], and identifying how the GDPR could be enforced can help increase data controller compliance while protecting data subjects' personal data [166].

From a technical perspective, PETs have been adopted in response to the need for digital tools to protect personal data and assess privacy risks. PETs are a "disruptive set of technologies and approaches which, when combined with changes in wider policy and business frameworks, could enable the sharing and use of data in a privacy-preserving manner. They also have the potential to reshape the data economy and to change the trust relationships between citizens, governments, and companies" [236]. In realising the potential for PETs, the area of human-data interaction (HDI) has also been developed to recognise the importance of allowing data subjects to receive value from their data. HDI is a conceptual framework for ethical systems design, focused on systems that collect and process personal data [122]. HDI supports a move away from the use of personal data in ways that benefit centralising governmental or corporate structures much more than the people that the data describes. The key principles of HDI include:

- **Legibility:** Premised on the recognition that interactions with data flows and data processes are often opaque, legibility is concerned with making data and analytic algorithms both transparent and comprehensible to users.
- **Agency:** The means to manage "our" data and access to it, agency enables us to act effectively in these systems, as and when we see fit. This not only includes the ability to opt-in or opt-out of data collection and processing but also the broader ability to engage with data collection, storage and use, and to understand and modify data and the inferences drawn from it.

  - Lack of agency also leading to lack of transparency, where *how* to exercise rights of some sort is also unclear or even further hidden through e.g. dark transparency problems.

- **Negotiability:** The means to navigate data's social aspects, negotiability supports interaction between other data subjects and their policies. This enables the ongoing engagement of users so that they can withdraw from data processing either completely or in part, and can derive value from data harvesting for themselves.

By integrating human-centred design within the development process, data protection is considered prior to data being collected, placing data protection responsibilities on data controllers and limiting the need for enforcement of remedies by practically applying DPbD principles.

In this section, we outlined the legal and technological context to which data protection challenges can be resolved. We note that there have been supplementary guidance and developments that more directly address the responsibilisation of the personal data process and the potential harms caused by increased data collection, analysis, and sharing. While these solutions offer *ex ante* means of protecting personal data and introduce ethical and interdisciplinary considerations for data protection, the data subject is still only passively considered and not fully participating in the development of these practices. Given that data subject engagement was seen as beneficial for ongoing digital developments such as the implementation of online learning from Chapter 2.4.3, supporting the co-creation of data protection solutions with data subjects and other stakeholders may be considered. Additionally, while there is recognition that it could be beneficial for data subjects to have access to the value generated from their personal data, it is unclear how this value will be delivered to them as there is no framework for developing a common pool of resources to share the knowledge generated from the data, data subject experiences, and data protection resources. As these considerations of regarding privacy and data protection beyond individual rights and responsibilities are relatively novel, there has been little research done on how practical data protection solutions can be co-created and collaborated on. Later in this thesis, we will tackle these outstanding issues.

The next section explores specific data governance and data stewardship frameworks that bring the developments mentioned in this section to life.

3.2 Addressing the responsibilisation of data protection

In Chapter 2.3, we noted that many existing legal and technological solutions responsibilises the data protection process and relies on individuals to exercise their own data rights. We also highlighted how the responsibilisation of data protection manifests through case studies in Chapter 2.4. In this section, we discuss the state of the art in addressing this issue in more detail, and consider the applicability of these methods to protecting personal data and support co-created and collaborative data protection solutions.

As discussed earlier in Chapter 2.3.2, different stakeholders within the data protection process have different interests and relationships with each other and the data itself. As a result, data governance mechanisms that aim to protect personal data could take these stakeholder tensions into account. Within data governance considerations, data stewardship refers to the process by which "individuals or teams within data-holding organisations ... are empowered to proactively initiative, facilitate, and coordinate data" towards the public interest [234], embodying the responsible planning and management of common resources [7]. Data stewards may facilitate collaboration to unlock the value of data, protect actors from harms caused by data sharing, and monitor users to ensure that their data use is appropriate and can generate data insights. New data stewardship frameworks, such as data trusts, data foundations, and data cooperatives have been devised to protect data subjects as well as involve them and other stakeholders in the co-creation of data protection solutions [5] [48]. While these data stewardship frameworks may help mobilise data protection rights, there are significant organisational, legal, and technical differences between them. Furthermore, they also face definitional, design, and data rights-based challenges. The benefits and challenges faced by these frameworks in context of data subject engagement for protecting personal data are discussed in this section and summarised in Table 3.1.

A data trust is a legal structure that facilitates the storage and sharing of data through a repeatable framework of terms and mechanisms so that independent, fiduciary stewardship of data is provided [106]. Data trusts aim to increase an individual's ability to exercise their data protection rights, empower individuals and groups in the digital environment, actively define terms of data use, and

Table 3.1: Data trust, data foundation, data cooperative, and data collaborative stewardship models summarised by their benefits and limitations in considering data subject engagement.

Data Stewardship Model	*Benefits*	*Limitations*
Data trust	Uses trust law as a basis for providing independent, fiduciary stewardship of data [106] with "bottom-up" inclusion of data subjects [56] and helps align trust and trustworthiness between them and other stakeholders [168].	There are still specific operational strategy questions that need to be answered for the deployment of data trusts [235]. While there are examples of "bottom-up" data stewardship, there are no current tested examples [232] of data trusts specifically that demonstrate data subject engagement.
Data foundation	Provides a good governance model to minimise the risks of personal data breaches and other non-compliant data-related activities by building data usage, sharing and re-usage environments that are trustworthy-by-design [218].	Beneficiaries are not required within the model and even if they are included, data subjects have limited rights in a foundation compared to a trust [191], with limited opportunities for direct engagement.
Data cooperative	Reduces the responsibilisation of the data protection process for data subjects through jointly pursuing collective interests [3] and can use data rights to advocate for data subjects on their behalf [179].	Data subjects may not be able to act independently from the group given the cooperative's group aims. Requiring contract or incorporation to establish rights and obligations could also reintroduce collaboration, engagement, and mobilisation challenges [3].
Data collaborative	Harnesses privately held data towards the public good with distinct goals to solve societal problems through collaboration between organisations from diverse sectors [253].	Data collaboratives focus on exchanging data by initially mitigating risks and harms to individuals and communities [254] but it is unclear how individual data subjects can directly engage.

support data use in ways that reflect shifting understandings of social value and changing technological capabilities [4]. Although data trusts refer to trusts in the legal sense, they also imply a level of trustworthy behaviour between data subjects and other data trust stakeholders [168]. While data trusts are promising in their ability to use trust law to protect data rights, it is currently unclear what powers a trustee tasked with stewarding those rights may have, and the

advantages the data subjects as the trust's beneficiaries (the person or persons who are entitled to the benefit of any trust law arrangement) may gain [4]. Data trusts could in theory support responses to certain data subject rights requests, particularly through access requests, but it may be difficult to benefit from other rights such as portability and erasure to support data subjects through trusts. In the latter case, there may be tensions regarding trade secrets and intellectual property [56] [4]. Moreover, the agency that data subjects may exercise within the data trust mechanism remains an open question. Efforts have encouraged the creation of "bottom-up" data trusts that aim to empower data subjects to control their data. While data subject vulnerability and their limited ability to engage with the day-to-day choices underlying data governance is acknowledged [56], many data trusts remain top down in nature and overlook the data subject's perspective.

It is still unclear how existing fiduciary structures in data trusts can fully realise their fiduciary responsibilities towards data subjects within digital spaces [151] [49] [232]. Previous pilots have attempted to clarify how data trusts can be put into practice (although without the application of trust law) by supporting the initiation and use of data trusts with a data trust life cycle [169]. Recent projects such as the Data Trusts Initiative can help clarify how the bottom-up data trusts model can operate in practice in realising fiduciary responsibilities [50]. Other frameworks, including data foundations, data cooperatives and data collaboratives, have also included citizen representation and engagement as an integral part of their design [126]. There are, however, still many practical challenges in this respect [235], particularly with questions relating to scaling and sustaining data sharing [137]. To address some of these challenges, data foundations have been developed as a good governance model for "responsible and sustainable non-personal and personal data usage, sharing, and re-usage by means of independent data stewardship" [216]. Data foundations rely on foundation law and view data subjects as potential beneficiaries within the model [218]. However, beneficiaries are not required within the model and even if they are included, data subjects have limited rights in a foundation compared to a trust [191].

A data cooperative is a group that perceives itself as having collective interests, which it would be better to pursue jointly than individually [3]. Cooperatives are "autonomous associations of persons united voluntarily to meet their common economic, social, and cultural needs and aspirations through a jointly-owned

and democratically-controlled enterprise" [125]. Data cooperatives can harness the value of data in common, where the growing real-time ubiquity of digital information could help its members plan more justly and efficiently than the price mechanism in our data-driven economy [160]. For example, the US-based Driver's Seat Cooperative is a driver-owned data cooperative that helps gig-economy workers gain access to work-related smartphone data and get insights from it with the aim limiting power imbalances between drivers and platform companies in the gig economy [209].

Data cooperatives can liberate personal data through data subject access requests and can advocate for data subjects on their behalf [179]. However, cooperatives often rely on contract or incorporation to establish rights, obligations, and governance, which could reintroduce some challenges related to collaboration and mobilisation the framework was intended to limit [3]. There may also be tension between reconciling individual and collective interests [47]. Navigating conflict in cooperatives may be carried out through voting or other governance structures, where data cooperatives may function as fiduciaries as well [121]. Similarly, data collaboratives [233] can harness privately held data towards the public good through collaboration between different sectors. Data collaboratives differ from other frameworks such as data trusts because the former have the distinct goal to solve societal problems through collaboration between organisations from diverse sectors [253]. They can support the rethinking of rights and obligations in data stewardship [252] to mitigate inequalities and data asymmetries [272].

3.2.1 Data stewardship challenges

Despite many efforts to help define and clarify the legal, organisational, and technical dimensions of data stewardship frameworks, one of the challenges is that no broadly accepted definition of data stewardship has emerged [217]. Their broad applications and widespread theoretical adoption have resulted in varied definitions and so require further disambiguation from each other to implement [224]. Even if these frameworks are clearly defined, data trusts and data collaboratives rely on separate legal structures to facilitate the protection of personal data through the creation of a new data institution using legal means [171]. It is acknowledged that each of these frameworks have the legal safeguarding of data subjects at their core. Nonetheless, the requirement of additional legal structures could further complicate the data protection process for both the

organisations willing to adopt these frameworks as well as data subjects' ability to engage with them.

Data stewardship frameworks also face several design challenges associated with the inclusion of data protection principles and data subject engagement. Although DPbD may be considered [217], the frameworks may still focus more on how the data generated can be used for specific purposes given their limited reference to data subjects' rights and agency over their personal data. While bottom-up approaches focusing on data subject agency are increasingly being considered as integral to the creation of existing data stewardship frameworks, they are not mandatory and may differ in their application. Although data subjects can be both settlors (a person who settles property on trust law for the benefit of beneficiaries) and beneficiaries within data trusts and beneficiary members in data cooperatives, individuals and groups of data subjects may still be excluded from participation in two circumstances: firstly, where they have not been consulted in the design of the framework and secondly, where there is a lack of clarity on what the a bottom-up approach entails [232] with uncertainty over how genuine and appropriate engagement mechanisms can be deployed [4]. Moreover, it is unclear whether or how existing data stewardship mechanisms apply participatory and action research-based solutions [22] to ensure that data subjects' preferences and perspectives are substantively taken into account as part of ongoing governance [196].

Finally, data-related rights may not be fully realised within current data stewardship frameworks. Although data stewardship frameworks benefit from not requiring extra legislative intervention that can take time to produce and is difficult to change [4], the frameworks also do not always interface with existing public regulatory bodies and their mechanisms which enforce data-related rights as part of their solution. It is also unclear how current data stewardship frameworks would support data subject recourse should there be personal data breaches. Data cooperatives often do not preserve privacy as a first priority [3]. While data trusts may introduce trustees and experts that are able to prevent potential data-related harms [4], it is not compulsory for them to do so. Given that seeking remedies from data protection harms is not mandatory within existing data stewardship models, data subjects may be left with limited support on how to exercise data subject rights under data protection regulations.

In this section, we introduced and summarised the different data stewardship

frameworks that could support greater protection of data subjects' personal data. With the recognition that data protection could be addressed beyond data protection law by involving other forms of regulation and socio-technical frameworks, the data stewardship frameworks as outlined in this section focus on supporting greater protection of data subjects' personal data. However, direct and ongoing data subject engagement and participation have not been fully considered within these data stewardship mechanisms. Crucially, the proposed data stewardship frameworks do not facilitate the common and collaborative pooling of resources as was identified to be useful for data protection in Chapter 2.4. Given the importance of power imbalances, we were surprised that different stakeholder tensions, outlined in Chapter 2.3.2, are not directly addressed in these data stewardship frameworks as there is little recognition of stakeholder roles outside of those outlined within data protection regulation. Additionally, our concern is that even with the focus on "data" within these stewardship frameworks, the application of data protection rights and principles as we explored in Chapter 2.2 may be limited given the lack of focus on data subjects in supporting the protection of their personal data, the exercise of their rights, and recourse for data harms.

Noting the importance of involving data subjects in the data protection process through co-creating and collaborating on data protection solutions through pooling common resources, we are interested in assessing whether and how data protection and data have been addressed in a commons. To do so, in the next section, we outline how considering data protection through the lens of a commons could tackle these challenges.

3.3 Co-creating collaborative data protection solutions

In Chapter 2.3.4, we noted that stakeholder tensions complicate the data protection process, where different stakeholders have different aims and objectives when it comes to the collection, processing, and sharing of personal data. Current data stewardship solutions may not directly address this through collective means, meaning that data subjects may be unable to co-create data protection solutions. In this section, we examine how the commons addresses these tensions and how it encourages collaboration and co-operation to solve these issues.

The commons, as developed by Elinor Ostrom, considers individual and group collective action, trust, and cooperation [174]. The commons guards a common-pool resource (CPR), a resource system that is sufficiently large as to make it costly to exclude potential beneficiaries from obtaining benefits from its use and may be over-exploited. Respecting the competitive relationships that may exist when managing a CPR, the commons depends on human activities and CPR management follows the norms and rules of the community autonomously [174]. The CPR enables "transparency, accountability, citizen participation, and management effectiveness" where "each stakeholder has an equal interest" [112]. Central to governing the commons is recognising polycentricity, a complex form of governance with multiple centres of decision-making, each of which operates with some degree of autonomy [178]. Its success relies on stakeholders entering contractual and cooperative undertakings or having recourse to central mechanisms to resolve conflicts [176]. The norms created by the commons are bottom-up, focusing on the needs and wants of the community and collectively discussing the best way to address any issues. This is illustrated by Ostrom's case studies of Nepalese irrigation systems, Indonesian fisheries, and Japanese mountains. These commons structures have enabled communities to find stable and effective ways to define CPR boundaries, define the rules for its use, and effectively enforce those rules [177].

From these case studies, Ostrom identifies eight design principles that mark a common's success with a robust, long-enduring, CPR institution [174]:

1. **Clearly defined boundaries**: Individuals or households who have rights to withdraw resource units from the CPR must be clearly defined, as must the boundaries of the CPR itself;

2. **Congruence between appropriation and provision rules and local conditions**: Appropriation rules restricting time, place, technology, and/or quantity of resource units are related to local cognitions and to provision rules requiring labour, material, and/or money;

3. **Collective-choice arrangement**: Most individuals affected by the operational rules can participate in modifying the operational rules;

4. **Monitoring**: Monitors, who actively audit CPR conditions and appropriate behaviour, are accountable to the appropriators or are the appropriators;

5. **Graduated sanctions**: Appropriators who violate operational rules are likely to be given assessed graduated sanctions (depending on the seriousness and context of the offence), from other appropriators, by officials accountable to these appropriators, or by both;

6. **Conflict-resolution mechanisms**: Appropriators and their officials have rapid access to low-cost local arenas to resolve conflicts among appropriators or between appropriators and officials;

7. **Minimal recognition of rights to organise**: The rights of appropriators to devise their own institutions are not challenged by external governmental authorities; and

8. **For larger systems, nested enterprises for common-pool resources**: Appropriation, provision, monitoring, enforcement, conflict resolution, and governance activities are organised in multiple layers of nested enterprises.

As the commons on its own focuses on creating a framework to be adapted to different cases and environments, we explore how these principles have been applied to digital- and data-related settings, and how might it be applied to data protection more specifically. In this section, we examine how the knowledge and information commons, data commons, and urban commons have considered data and data protection practices within their respective commons frameworks.

3.3.1 Knowledge and information commons

To address the rise of distributed, digital information, Hess and Ostrom developed the information or knowledge commons, where knowledge is the CPR [113]. As new technologies enable the capture of information, the knowledge commons recognises that information is no longer a free and open public good. Instead, it now needs to be managed, monitored, and protected for archival sustainability and accessibility. Crucially, the commons addresses data-related governance challenges that arise due to spillovers created by the reuse of data, so increasing its value over time [44]. This is further exemplified when data is linked together, creating new uses and value for the same data. Without a commons, the newly generated knowledge may not be available to the original creators of the data in the first place. As a result, the knowledge commons can support data subjects in accessing the personal and social value of their data while ensuring its quality and

storing it securely.

In assessing the feasibility of a knowledge commons, Ostrom's Institutional Analysis and Development (IAD) framework can be used to study an institution's community, resource dynamics, and stakeholder interests. The IAD supports the creation of a commons and analyses the dynamic situations where individuals develop new norms, rules, and physical technologies. The IAD framework is adaptive as the analysis helps us think about economic structures that promote collective ownership which is economically, ecologically, and socially beneficial [257]. Adopting the IAD framework's core sections on biophysical characteristics, action arena, and overall outcomes, the framework acts as a "diagnostic tool" that investigates any subject where "humans repeatedly interact within rules and norms that guide their choice of strategies and behaviours" [113]. Institutions are defined as formal and informal rules that are understood and used by a community. Central to the IAD framework is the question "How do fallible humans come together, create communities and organisations, and make decisions and rules to sustain a resource or achieve a desired outcome?". Broken down into three core sections, a knowledge commons can be assessed by its resource characteristics (the biophysical-technical characteristics, community, and rules-in-use), action arena (institutional changes and the process of voluntary submitting artefacts), and overall outcomes. Specifically for a knowledge or information commons, the IAD framework is useful because it supports investigation into how resources are actually governed and structures the empirical inquiry to facilitate comparisons, while avoiding unwarranted assumptions related to particular theories or models [222]. As part of the IAD framework, Ostrom identified seven rules by which institutions could be analysed [175]:

1. **Position**: The number of possible "positions" actors in the action situation can assume (in terms of formal positions these might be better described as job roles, while for informal positions these might rather be social roles of some capacity);

2. **Boundary**: Characteristics participants must have to be able to access a particular position;

3. **Choice**: The action capacity ascribed to a particular position;

4. **Aggregation**: Any rules relating to how interactions between participants within the action situation accumulate to final outcomes (voting schemes etc.);

5. **Information**: The types and kinds of information and information channels available to participants in their respective positions;

6. **Pay-off**: The likely rewards or punishments for participating in the action situation; and

7. **Scope**: Any criteria or requirements that exist for the final outcomes from the action situation.

In advancing the practical application of the IAD framework into new use cases, the framework has been adapted to create building blocks for developing a commons. Ostrom adapted her design principles into key questions to create actionable means for problem solving [175]. Translating the IAD framework's core sections of biophysical characteristics, action arena, and overall outcomes, McGinnis transposes these questions, abstract concepts, and analytical tools to a detailed study of specific policy problems or concerns [152]. For example, concerns related to information flows, the enforcement of rules, and expressing outcomes as dynamic processes help address power and information asymmetries between different stakeholders. McGinnis encouraged users of the framework to adapt them in ways that best suit their applications to the factors deemed most important for understanding the research puzzle or policy concern that serves as the focus of researchers' own work. A summary of McGinnis' steps of analysis are:

1. Decide if your primary concern is explaining a puzzle or policy analysis.

2. Summarise two to three plausible alternative explanations for why this outcome occurs, or why your preferred outcome has not been realised; express each explanation as a dynamic process.

3. Identify the focal action situation(s), the one (or a few) arena(s) of interaction that you consider to be most critical in one or more of these alternative explanations.

4. Systematically examine categories of the IAD framework to identify and highlight the most critical.

5. Follow the information flow in each of these focal action situations.

6. Locate adjacent action situations that determine the contextual categories of the focal action situation. This includes: outcomes of adjacent situations in

which collective actors are constructed and individual incentives shaped, rules are written and collective procedures established, norms are internalised and other community attributes are determined, goods are produced and inputs for production are extracted from resource systems (that may need replenishment), and where evaluation, learning, and feedback processes occur.

7. Compare and contrast the ways these linked and nested action situations are interrelated in the processes emphasised by each of your alternative explanations.

8. Identify the most critical steps for more detailed analysis, by isolating components of adjacent action situations that determine the context currently in place in the focal action situation(s), and that if changed would result in fundamental changes in outcomes.

9. Draw upon principles of research design or evaluative research to select cases for further analysis by whatever methods are best suited to that purpose.

When creating a knowledge commons, Strandburg [222] also mapped the IAD framework into research questions as a means to support the planning and governing process of a commons. This includes the interview process for gathering participants and turning those interviews into practical goals and objectives for commons governance. The knowledge commons has also been applied to privacy through the Governing Knowledge Commons framework [204] by conceptualising privacy as information flow rules-in-use constructed within a commons governance arrangement and by considering Nissenbaum's approach to privacy as contextual integrity [164].

An example of an information or knowledge commons is a university repository with academic and research data resources [113]. Developing a university repository requires multiple layers of collective action and coordination as well as a common language and shared information and expertise. The local community, academics, and researchers, can contribute to the repository as the more it is used, the more efficient the use of resources is to the university as a public institution. Others outside that community can browse, search, read, and download the repository, further enhancing the quality of the resource by using it. By breaking down large, complex, collective action problems into action spaces through the IAD framework and using Ostrom's design principles for governing a commons, institutions and organisations can more accurately meet the needs of those in the

community, including how information, knowledge, and data can be used to serve the common good.

From a technological perspective, the open source and open software communities can also be seen as knowledge commons, where software are freely and publicly available for commercial and non-commercial uses. The software tools are also openly developed and anyone is able to contribute. Organisations such as The Open Usage Commons [172] help project maintainers and open source consumers have peace of mind that projects will be free and fair to use. Platforms such as Wikipedia and the Wikimedia Commons [263] are public domain, freely licensed resource repositories that are open and can be used by anyone, anywhere, for any purpose, and have been demonstrated to conform to Ostrom's design principles [83]. Adapting the knowledge commons to participatory sensing, Macbeth found that sustainable, collective, self-organising management of the information and knowledge in participatory-sensing applications is both possible and competitive when Ostrom's IAD framework is applied [141].

Ostrom's design principles and the IAD framework can limit the responsibilisation of the data protection process because they encourage active engagement of data subjects and consider how data can be protected through the development process while increasing its value. The analysis steps, questions, and framework encourages iterative means of creating a commons and supporting the co-creation process. The IAD framework recognises that the expectations, possibilities, and scope of information and data can be different as more knowledge is included within the commons. These principles are also useful in considering data protection solutions because they recognise that there is no one-size-fits-all fix, supporting more flexible and adaptable ways of achieving the commons' goals. Incorporating existing regulations and policies into the commons for data protection allows data subjects to find specific solutions to their challenges. The commons can support them by developing a more holistic understanding of the data protection landscape of the specific domain, collaborating with other data subjects or stakeholders to co-create individual data protection preferences, and exercising their data protection rights with the support of the community that has been harmed.

3.3.2 Data commons

Ostrom's commons framework has also been applied to data commons which guard data as a CPR. Traditionally, such data commons focus on data distribution and sharing rather than data protection [81]. Research data commons such as the Australian Research Data Commons (ARDC) [10], the Genomic Data Commons (GDC) [159], and the European Open Science Cloud (EOSC) [69] all attempt to further open science and open access initiatives. The ARDC is a government initiative that merges existing infrastructures to connect digital objects and increases the accessibility of research data. The National Cancer Institute also has a GDC that is used to accelerate research and discovery by sharing biomedical data using cloud-based platforms. With a research-oriented focus, the GDC does not house or distribute electronic health records or data it considers to be personally identifiable but still had safeguards against attempts to re-identify research subjects [129]. Other examples include the Data Biosphere [58] that uses a modular, community-driven, open, standards-based governance and the National Library of Medicine that requires the necessity for security, searchability, standardisation of metadata, and the management of access control [264]. In Europe, the EOSC is a digital infrastructure set up by the European Commission for research across the EU, with the aim to simplify the funding channels between projects. The EOSC was inspired by the F.A.I.R. principles, representing Findable, Accessible, Interoperable and Reusable data sharing and aims to become a "global structure, where as a result of the right standardisation, data repositories with relevant data can be used by scientists and others to benefit mankind" [69]. While these frameworks recognise that the information and knowledge included are collectively created, their implementations are hierarchical and top-down as they were created through structured committees, serving as a data repository platform that enables research reproducibility [102]. As a result, they may have limited input from archive participants, repository managers, or public consultation processes and do not take Ostrom's principles into account. Additionally, given the goals and objectives of these commons, by nature, they prioritise data sharing, data curation, and reuse, over data protection. The EU frameworks acknowledge the GDPR as the source for regulatory data protection. However, it is currently unclear as to how this is implemented. While these data commons can be fruitful for furthering research and opening up data for reuse, they do not take into consideration the data subjects that created the data in the first place, as much of the data stored in these commons are not considered personally identifiable information. As a result,

existing data commons alone are insufficient for protecting personal data as they are designed without data subjects' personal data in mind.

A data commons is useful for managing the pooling of common resources because mechanisms such as licensing for data archives may not be useful for data protection even if they limit forms of data reuse [103]. For example, in a commons created for data archives, the data subject can maintain more control over their data through the research process [189], reducing the risks of personal data being misused. A data or knowledge commons can also be useful for considering the intellectual property rights of mass-participation content creation on social networking sites and in pervasive computing, where the commons could support the use of collective intelligence and knowledge sharing to address systemic problems which threaten the sustainability of institutions and physical infrastructures [184]. A data commons can help distribute the benefits of data as a resource widely and equitably, without commodifying or privatising it [180]. The commons has also been considered for governing emerging technologies as the commons can help mitigate individual and collective risk [219]. Beyond personal data, raw non-personal data may also be governed by the commons to manage data without data ownership [80]. This is especially important when curated data that used to be in the public domain no longer are, with wider ramifications if the data is socially and politically sensitive, such as Twitter data on the 2014 Hong Kong Umbrella Movement pro-democracy protests [241]. While digital data archives aim to preserve, reuse, and promote ethically sound, methodologically well-grounded research, there continues to be insecurity by researchers about data sharing, where social media data sharing may become hidden and informal [261]. A data commons created with data protection as applied to archives of curated data could also help clarify who the data controllers are from a wide range of archive owners, dataset owners, or participants, and identify who is accountable to and for the publicised data and the resulting reuse outputs. Without a data commons, questions such as "Who maintains control over curated data?", "How can data controllers limit who and how collected data is reused?", and "How can data subjects exercise their data protection rights when sensitive and identifiable personal data that could potentially be de-anonymised is curated?" remain difficult to answer.

3.3.3 Urban commons

In addition to a knowledge commons and data commons, urban commons frameworks have also considered data protection principles. An urban commons represents shared material, immaterial, or digital goods in an urban setting [78]. Urban commons or data commons frameworks applied to urban environments have been created in an attempt for governments to take more responsibility over their citizens' personal data [68]. An urban commons represents resources in the city that are managed by its residents in a non-profit oriented and pro-social way [57]. It is a physical and digital environment that aims to more appropriately utilise an urban space for public good, formed through a participatory, collaborative process. To tackle surveillance and monitoring challenges in urban settings, co-design processes could be considered to create ethical monitoring solutions as well as supporting collective representation to negotiate such surveillance practices [40].

Urban commons aim to increase the transparency of how city data is used and provide accountability should users and data subjects want their data withdrawn. For example, the European projects DECODE [68] and the gE.CO Living Lab [88] both encourage citizens to be part of a collaborative process in creating communal urban environments that more accurately represent the community. The DECODE data commons project "provides tools that put individuals in control of whether they keep their personal information private or share it for the public good" [68] with the focus on city data in four different communities. The project not only created an application to support user control over their data [55] but also produced documents for public use on community engagement, citizen-led data governance, and smart contracts to be applied to urban environments. The outcomes from the project have been applied to local European projects such as Decidim in Barcelona to create open spaces for democratic participation for cities and organisation through free, open-source digital infrastructures [54]. Further, the DECODE project continues to shape the EU's direction when it comes to policy-making for digital sovereignty [28]. The gE.CO Living Lab created "a platform for bringing together and supporting formal groups or informal communities of citizens" who manage co-creation spaces and social centres created in regenerated urban voids [88]. The Lab's aim is to foster "sharing and collaboration between citizens and establish a new partnership between public institutions and local communities, setting forth new models of governance of

the urban dimension based on solidarity, inclusion, participation, economic and environmental sustainability" [88]. In the UK, the Bristol Approach also aimed to explore the potential of a commons as a tool for social change in an urban context through citizen sensing [230]. The Bristol Approach was seen as a way to involve the local government not just in producing data for citizens to use but also in responding to citizens' needs as presented by new technologies [190]. While the project was successful in mobilising resources for discussion, self-representation, and action, there were still community tensions related to who was able to participate in public engagement. As cities become more digitally connected and more data is being collected from its citizens, an urban commons increasingly focuses on data both in determining how information and resources can be created and shared within a community as well as focusing on citizens' personal data.

In this section, we have looked at the different data-related commons frameworks that can support co-created and collaborative approaches to protecting personal data. We illustrate how different commons have been useful for supporting different stakeholder considerations of data protection. Using Ostrom's design principles and polycentricity as a form of governance, the commons framework has been adapted for information, data, and urban environments. However, these commons may not fully engage data subjects in co-creating data protection solutions to maximise the value of their data or support data subject recourse through the exercise of data subject rights. In the next section, we examine how these principles and frameworks can be adapted for data protection to address data-related harms, as discussed in Chapter 2, by recognising the limitations of both laws and technologies, encouraging collaborative solutions for protecting personal data, and allowing data subjects to regain autonomy of their data protection process.

3.4 A Commons for Data Protection

More recently, organisations focusing on data governance and data stewardship have explored the use of a commons for data with applications specifically focused on data protection. The Ada Lovelace Institute has identified a data commons as a means to tackle data-related issues such as consent and privacy by mapping Ostrom's principles to specific GDPR principles and articles [183]. The focus on creating a commons for data protection draws attention to the sharing and

dissemination of information and expertise as it relates to data, encouraging a more open and collaborative environment. By sharing the data protection information that is available, responsibilisation can be limited, where resources are pooled for collaborative decision-making instead of individuals having to understand everything on their own. This can minimise the impact of data-related harms as a preventative method rather than a reactive one. Interdisciplinary research directions for establishing commons governance and participation in a privacy commons have also been identified [155]. In developing the practical basis for developing new forms of data stewardship through a commons, the Ada Lovelace Institute has also compiled a list of commons projects, mapping them to Ostrom's principles and creating a set of design principles for data stewardship [6]. More broadly looking at the value of data, the Bennett Institute and Open Data Institute (ODI) have mapped Ostrom's principles to examples of how our data is used in a data-driven economy, highlighting the need to "provide models for sharing data that increase its use, capture positive externalities and limit negative ones so we can maximise the value of data to society" as well as include trustworthy institutions that together govern who can access what data "in accordance with the social and legal permissions they are given" [21]. As a result, considering data protection as an essential part of the commons can support the redistribution of data's value to help rebalance power between data subjects and data controllers.

Beyond the sharing and protection of personal data, a commons for data protection can act as a consensus conference [8] to encourage dialogue among data subjects, experts, and policy-makers, experts and ordinary citizens, creating new knowledge together for the common good. Van Laerhoven and Barnes consider how community development efforts can target collective action potential and self-governance capacity of commoners by combining the commons work of community development theorists and practitioners [250]. While existing data stewardship frameworks may take its members vested interests into consideration, the commons has a number of advantages for data subject engagement. First, a commons, through its stakeholder considerations and bottom-up norms, can directly engage data subjects in the creation and iterative improvement of the framework. Data subjects are then able to actively and continuously reflect on their individual and community preferences when it comes to managing a CPR. Commons principles have been considered for enriching and protecting data-enabled social knowledge through cooperativist production [105]. Second, it can advance the protection of personal data as part of democratic and participatory

governance [87]. Privacy and data protection may be addressed directly not only as legal rights but also as part of the political, social, and cultural landscape [63]. Third, it can offer an alternative form of data stewardship in that it applies polycentric design principles [63] and because these principles adopt public engagement methodologies to engage with and empower data subjects. These methodologies, which have their roots in Human-Computer Interaction and STS, can increase public engagement not only with science, but also with legal, policy, and technical innovations [267] [266] [220]. Public engagement beyond the development of science, law, and policy is also necessary for establishing trust [271]. The commons for data protection can support direct engagement between data subjects and to other stakeholders through its infrastructure as well as the application of conflict-resolution mechanisms based on Ostrom's design principles. The focus of these methods on worst-case scenarios, such as data breaches or privacy violations, is particularly helpful for data subjects [240]. When addressing the likelihood of these risks occurring, the collective identification of shared goals and purposes can be enabled. Data subjects can collectively exercise their data subject rights and co-create data protection solutions that both benefit them individually and as a group. Data subject agency, engagement, and empowerment may thus be garnered through the democratic expression of individual preferences towards improving individual and collective commitment towards a shared data protection goal. At the same time, a commons for data protection can help carefully juggle the interdependence between civil society and legal-political mechanisms [53]. A commons can help limit power asymmetries between data subjects and data controllers if structured attention to who, how, and what is excluded from digital data and information commons, to what effects, and how we can prevent undue and harmful exclusion are considered [192].

From a technical perspective, Diaconescu and Pitt identify the need to build "pro-social socio-technical systems" to balance transparency and privacy, where identified pathologies stem "from regulatory choices and associated power struggles" [60]. A commons-based peer production model has also been developed to codify values such as community building, objective accomplishment, monetary value, social value, and reputation measures from web analytics indicators [157]. This could be useful for a commons focused on data protection as a commons' polycentricity directly engages with the stakeholder tensions outlined in Chapter 2.3.2. Even within open systems, adapting Ostrom's common-pool resource management methodology into a formal system of retributive justice in self-

organising electronic institutions helps to reduce costs, marginalise non-compliant behaviour, and increase collective utility [185].

While data protection has been considered as part of the commons process as identified in Chapter 3.3, including data subjects and their communities is not a requirement when considering how their personal data can be protected. Creating a data protection-focused data commons could help identify how much understanding and control data subjects have over their personal data and support them in choosing their data protection preferences. A commons for data protection does not require the creation of a new legal framework, but rather, operates within the current data infrastructures used by data subjects and acknowledges the limitations of existing laws, technologies, and policies that steward data. Thus, the focus on data protection as part of the data commons shifts data protection responsibilities away from the individual alone and to communities, where knowledge, expertise, and experiences can be pooled together to identify working solutions. Although personal data is still kept personal and private, the collaborative nature of sharing, discussion, and advising on data protection problems opens up potential options for everyone to support informed decision-making and achieve data protection preferences through a data commons. Given that data-based systems do not exist distinct from the ethical qualities of the social milieu in which data-intensive technologies are produced [247], a bottom-up commons for data protection can not only help reduce the data-related ethical uncertainty and contingency in these spaces but also support negotiations between stakeholders within this social space.

In this section, we outlined how the theoretical applications of data protection to a commons has been more recently explored. While data protection has been considered in a commons, the creation of a socio-technical data protection-focused data commons to support data subjects in the co-creation of data protection solutions has yet to be established. Drawing upon interdisciplinary literature, we explored how creating a commons specifically for data protection can directly support data subjects' rights and limit the responsibilisation of the data protection process through polycentricity, engagement, and collaboration.

3.5 Summary

In this chapter, we have surveyed the recent literature proposing solutions for data protection and responsibilisation challenges, and identified some data stewardship solutions to resolve these issues. We note the following points:

- There are many proposed legal, technological, socio-economic, and policy solutions for data protection and responsibilisation challenges. However, many still depend on individuals to address their own data protection concerns.
- Collaborative solutions have been considered for resource management and protection, including for information and data, but so far none consider data protection as the main objective or data subjects as part of the co-creation process.
- While data protection has been included in collaborative solutions, existing commons may not focus on directly protecting data subjects' personal data or supporting the exercise of their rights in practice.
- Data commons have been considered for data protection in theory, but a framework for creating a data commons that focuses on data protection and centers the data subject has yet to be established.

In the rest of this thesis, we apply and further develop the existing theories and frameworks surrounding the creation of a data commons for data protection to create a data protection-focused data commons, which we have discussed in this chapter.

In the next chapter, we introduce the building blocks for establishing a data protection-focused data commons by conducting interviews with experts, which we will evaluate in a number of case studies in subsequent chapters. Informed by Ostrom's principles and theories on polycentricity, a commons focused on data protection enables the co-creation and collaboration on data protection solutions to support data subjects' personal data preferences.

Chapter Four

Improving existing commons for data protection

In Chapter 2.3, we noted the interdisciplinary data protection challenges with regards to stakeholder tensions, responsibilisation, and power imbalances between data subjects and the data controllers that collect, analyse, and share their personal data. Our assessment of the data stewardship landscape in Chapter 3.2 identified how a data protection-focused data commons could be created to prioritise data subjects' personal data preferences within the data protection process through engagement and co-creation.

In order to assess whether a data-protection focused data commons is appropriate as a socio-technical framework for data stewardship as posed at the beginning of this thesis, we need to understand how a commons has been implemented in practice, and whether the commons could be improved and adapted specifically for data protection. As demonstrated from our case studies in Chapter 2.4, the challenges resulting from the increased networks of personal data moving within our data driven-society are particularly difficult to tackle. This is particularly due to the speed of technological advancements and reliance on data to power, inform, and innovate private and public infrastructures within our digital economy. Addressing our first research question on whether a data protection-focused data commons is appropriate as a socio-technical framework for data stewardship, we break down the question into action points that address the socio-technical aspects

for supporting data subjects' data protection process by asking: How do we make sure that data subjects fully understand how their data is used? To what extent can we mitigate potential data-related harms without detriment to innovation and technological development? What is the most effective way of involving data subjects in the data protection process, if at all? How can we generate greater communal and public value from the personal data that data subjects provide and allow them to express their personal data protection preferences simultaneously?

Before distilling these challenges into requirements by applying the IAD framework and steps of analysis identified in Chapter 3.3.1, we first need to understand how data protection has been considered within existing commons and whether a data commons for data protection could be adapted. To do this, we aim to find out more about how existing commons were created, what the associated challenges related to data protection were, and support the implementation on a commons. To investigate those aims, we conduct interviews with commons experts to identify the challenges of building a commons and important considerations for a commons' success. We established four primary interview questions to explore whether using data subject rights and data protection principles to support a data protection-focused data commons for collaborating and engaging data subjects in co-creating data protection solutions is suitable both in theory and in practice as a socio-technical means for data stewardship:

1. How, if at all, did interviewees work on identifying and solving data protection challenges?

2. How can the challenges of implementing a data commons best be overcome, specifically for data protection?

3. What do interviewees think could be done better in terms of creating a commons?

4. Is a commons framework useful for ensuring that personal data and privacy are better protected and preserved?

From the commons experts' contributions to our questions, we thematically analyse our findings. Using our results, we develop a practical framework for including data subjects in the data protection process through a data commons by adapting the IAD framework for assessing and implementing the data protection-focused data commons.

In this chapter, we make the following contributions:

- We conduct interviews with commons experts to identify the socio-technical, political, and institutional barriers to creating a commons and the challenges of incorporating wider data protection principles into a commons.
- We summarise our interview findings and identify key themes to enable the development of a data protection-focused data commons as a socio-technical framework for data stewardship.
- We adapt these themes to create an IAD framework for data protection to support the practical deployment of the commons, further identifying the next steps for establishing a data-protection focused data commons.

4.1 Method

In this chapter, we build upon our exploration of commons theories and analysis from Chapter 3 to assess how commons frameworks have been put into practice. As it is not evident that data protection is suitable to be governed within a commons, our interviews assess the extent to which the data protection challenges we outlined in Chapter 2.3 have been tackled within existing polycentric commons governance approaches. These include regulatory and compliance difficulties, addressing stakeholder tensions, and limiting the responsibilisation of the data protection process. The interviews conducted as part of the study act as a form of engagement that helps reveal the data-related power imbalances and suggested means to resolve them (Chapter 3.1). These solutions are driven by the Human-Computer Interaction and interdisciplinary engagement methodologies from Chapter 3.4.

We developed our study in three phases: identifying relevant commons and key informants, writing the interview questions, and conducting the interviews.

4.1.1 Identifying relevant commons and key informants

In Chapter 3.3, we identified different forms of commons frameworks that address the stakeholder tensions outlined in Chapter 2.3.4 and how a commons encourages collaboration and co-creation to solve these issues. For this study, we first identified which commons and experts to interview. In context of data protection, we

found that urban commons and data commons applied to urban areas are the most relevant for establishing a data protection-focused data commons because they represent a commons model that considers data protection and privacy. These commons are important for the development of a commons that focuses on data protection because they explicitly contrast smart city data governance and stewardship projects that increase the datafication and surveillance of urban environments. For example, Toronto's [33] and Portland's [43] Alphabet Sidewalk Labs smart city projects were scrapped due to concerns about the consolidation of data within big technology companies, lack of transparency about how public funds were to be used, and lack of public input during the development process. The Toronto Sidewalk Lab project relied on the creation of a Civic Data Trust, however, the Trust provided the potential to commercialise or licence access to the data once they had been collected, going against the public interest of that specific community [190]. As a result, urban commons environments that oppose the mass gathering of public data and engage with resource management "oriented towards use within the community, rather than exchange in the market" [215] most accurately represents how data protection is considered in a data protection-focused data commons.

The relevant commons identified for answering our research questions were found through conducting a literature review on recent self-described urban commons and data commons. The commons selected all used the commons to describe their work and their aims, with goals that emphasise co-creation and collaborative work with the community. As this thesis is written within the jurisdiction of the GDPR, an online search was conducted to identify European commons only.

Once the commons were identified, experts were chosen based on their expertise and experience in creating and developing an urban or data commons, and were contacted via e-mail. To ensure that we had a fair assessment of the commons development process, when contacting experts, we made sure that they had different levels of expertise, had different roles and responsibilities within commons development, and represented different communities. The size and scope of each commons project were also as varied as possible to gain a better understanding of how similar or different commons challenges may be throughout development.

Interviews were conducted to contextualise the role of the commons from different stakeholder perspectives and provided useful information into potential

challenges in the development process. Interviewees were told that this study contributes to our wider work on establishing a data protection-focused data commons to achieve data protection for data subjects in a collaborative way and allows them to co-create data protection policies with other data subjects and stakeholders.

Prior to the interview, key informants were given a participant information document and a consent form for them to sign and return. Once the interview was complete, a debrief was sent to the participant with more information about their data rights and our broader research.

4.1.2 Writing the interview questions

Key informant interview methods were used for writing the interview, with a semi-structured format to encourage discussion around the commons. The questions aimed to answer the research questions identified at the beginning of this chapter to explore the feasibility of a data protection-focused data commons. The interview questions are included in Appendix A.

Regarding the interview questions, the introduction set out the purpose of the interview, confirmed the interviewee's participation and consent to recording, and established a common understanding of a commons as well as outline their work of interest. Questions about the project aims and identifying problems asked the interviewee more about their experience on the project, including questions about their project team, the stakeholders with whom they interacted, and how they went about identifying potential problems throughout the project. Questions relating to problem-solving and overcoming challenges focused on both broader challenges relating to individual projects as well as ones that were about privacy and data protection specifically. If data protection or privacy were not explicitly mentioned by the interviewee, we asked whether and how these issues may have been considered. Overall project perspectives related to successes, limitations, and what interviewees might have done differently in relation to their commons project. Broader questions discussed considerations of what an ideal commons may look like. Before the interview ended, interviewees were provided with an opportunity to make any additional comments and ask questions. Finally, the summary section wrapped up the interview by providing more details about how this study fits into our wider body of research in relation to a data protection-focused data commons and gave interviewees another opportunity to ask questions.

In responding to the primary interview questions posed at the beginning of this chapter, we asked the experts those questions and supplemented them with the following questions, before engaging in further discussion:

1:

- How did you and the project team identify your project aims and what were some of the problems or challenges you considered during that process?
- What stakeholders did you interact with to solve some of these challenges?

2:

- How did you go about solving the challenges identified?
- Were there problems or challenges during the project that you didn't expect related to data?

3:

- What do you think are/were the successes of your role in the commons?
- How was this success achieved?
- What do you think are/were the limitations of your project as a commons?
- Is there anything you would do differently?

4:

- How do you think data and data protection can be best represented in the commons and in commoning?
- Can data subjects and participants' involvement in creating the commons support better data protection practices?
- What do you think is the ideal commons for data? Do you think it can be achieved? If so, how?

4.1.3 Conducting the interviews

Interviews were conducted either over the phone or conferencing software such as Skype, jit.si, or GoToMeeting based on the interviewees' preference. All interviews were conducted between March and November 2020 and lasted up to one hour. All interviews were recorded with the interviewee's consent. Once each interview was completed, audio recordings were placed into the MaxQDA qualitative data analysis software for immediate transcription and pseudonymisation. Once the transcription was finished, audio recordings were deleted.

4.2 Analysis and results

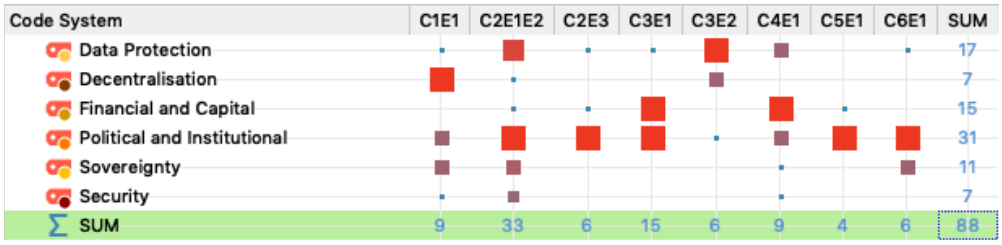
Nine experts across six commons were interviewed. The size, number of participants, and stakeholders varied across the commons, with three interviewees based in The Netherlands, two in the UK, one in Belgium, one in Germany, one in Italy, and one in Spain. Their roles and specialisms are listed in Table 4.1. The Reference *Cx* denotes the commons they contributed to and *Ex* denotes the expert. Role characterises the experts based on their responsibilities within the commons. Expertise describes their main contribution towards the commons.

Ref	*Role*	*Expertise*
C1E1	Academic	Privacy, Computer Science
C2E1	Technical	Privacy, Software Engineering
C2E2	Governance	Public Planning, Public Policy
C2E3	Policy	Commons Theory, Peer-to-Peer
C3E1	Policy	Technology, Public Research
C3E2	Academic	Privacy, Law, Information Science
C4E1	Policy	Third Sector, Community Engagement
C5E1	Policy	Community Development Planning, Public Research
C6E1	Research	Commons Policy, Community Engagement

Table 4.1: List of interviewees representing their commons project, role within the project, and their expertise.

Using MaxQDA, codes and tags were used to identify patterns for preliminary transcript analysis, identified in Figures 4.1 and 4.2. Although the interview centred around developing a commons and the challenges regarding data protection, human-centred themes, such as political and financial relationships, were mentioned the most. Based on these characterisations, main themes were drawn out and expanded upon from the interviews. For our interview transcript

analysis, we present our results in four sections: identifying data protection challenges, overcoming data protection challenges, improving the commons, and building a commons for data protection. We found that political and institutional barriers when it came to creating a commons were the most difficult to tackle, underlying how data and data protection are not necessarily seen as something that could be perceived as a commons. While data protection was discussed as part of the commons development process, there were limited applications to wider data protection principles such as those related to informing data subjects about their rights and the ability to exercise those rights against data controllers. All experts identified limitations within their own area of expertise, suggesting that these limitations, whether in law, technology, or policy, need to be identified within the commons to find more appropriate data protection solutions. Although the decision to use a commons was to provide certain levels of control and transparency of how data was collected, used, and processed, financial restrictions limited the potential impact of the commons framework and the extent to which a commons could scale. Interviewees further mentioned that working with stakeholders of different backgrounds helped everyone understand how a commons could be implemented and be beneficial for reaching data protection goals. Our interview findings are addressed thematically below by each primary interview question.



Code System	C1E1	C2E1E2	C2E3	C3E1	C3E2	C4E1	C5E1	C6E1	SUM
Data Protection									17
Decentralisation									7
Financial and Capital									15
Political and Institutional									31
Sovereignty									11
Security									7
SUM	9	33	6	15	6	9	4	6	88

Figure 4.1: Code matrix created from interview transcripts with all experts. Manually coded themes related to identifying problems and challenges were tagged and their frequencies are visualised based on how often they were discussed by interviewees. The most prominent challenges are those related to politics and institutions (31), followed by data protection (17), and financial and capital (15) related issues.

4.2.1 Identifying data protection challenges

From the interviews, the experts identified data and data protection challenges related to the commons based on their own role-specific experiences. Building upon this area of work from Chapter 3.3, these challenges include the difficulty

Figure 4.2: Code relation matrix created from interview transcripts with all experts. Manually coded themes related to identifying problems and challenges were tagged and their relationship with other themes are visualised based on how often they overlap, demonstrating how certain challenges are linked together. The most prominent relationships are political and institutional - financial and capital as well as political and institution - data protection. Data protection issues also demonstrate some overlap with other problems and challenges more generally.

establishing the scope of the commons regarding data and data protection, assessing how the commons can be beneficial to those who participate, and determining the value of data included as well as the data protection benefits.

First, in identifying the data protection challenges within commons projects, interviewees mentioned that the main aims of the commons were often provided by the project coordinators. The experts themselves only had partial input on the scope of the commons and how the commons was to be defined. An interviewee in a technical role said that following their core commons aim: *"The most important challenge there was to make it decentralised"* (C1E1). Another interviewee elaborated that: *"Essentially what [the coordinators] wanted was, they realised that this [issue] poses a threat to [users'] privacy and they wanted us to build a system from the same dataset"* (C2E2). However, it was clear to some experts that data and data protection challenges would only more clearly emerge once the foundation of the commons was established alongside other stakeholders due to the nature of commons building. One interviewee said: *"The project was, we have these technologies, we do not know how these are going to be because we haven't built it yet"* (C2E1). Another interviewee said: *"[One of the challenges is] striking a balance between openness and protection ... and then just institutionalising that with advanced information and communications technologies"* (C2E3). As a result, the precise scope of the commons needs to be flexible to accommodate changes during the development process and incorporate participant input. This includes being open to changes when it comes to how data is collected and managed within the commons itself.

When discussing the benefits of a commons to data subjects for protecting their personal data, there is a role of responsibility from experts to communicate the

options for protecting personal data: *"We played a role of coordination, and interaction with data subjects and data protection officers"* (C3E2). Another interviewee said that participants in a commons should understand that they have a real ability to have autonomy and sovereignty over their personal data, where the commons can support their preferences by operationalising this control: *"For the data commons to work, you need to be able to give citizens some kind of control over their data, and give them, some kind of like, choice of what the data was going to be used for or not used for"* (C1E1). Beyond the challenges laid out by project coordinators, interviewees also mentioned that there were data protection challenges that go beyond the practical creation of the commons and included theoretical, philosophical, and psychological aspects of people's relationship with privacy. One interviewee summarised this eloquently: *"So essentially three challenges: Money and difficulty in the social side, distributing the technology, and the philosophical 'who owns divulged data' in the community side"* (C2E1). Wider socio-economic issues surrounding data and internet access also need to be addressed when considering and implementing a commons framework: *"The first thing I became aware of is the inequality in our access to the internet"* (C4E1). One interviewee suggested that rather than considering the commons framework as something put on top of a community, think about a data commons as intrinsically part of community collaboration: *"Digital space is infinite, we can have an infinitely large number of people in it, but we are still biological beings, we are still constrained by our biology and our grey matter up here. We can still only really build closed connections to this relatively small number of people. The question is not, in my opinion, how can we make commons all over the place, but more how can we bring together this biological and digital realities to optimise what is happening"* (C6E1). While the commons itself may be valuable in terms of increasing accessibility to data and knowledge, it could be managed in a way that is accessible and easily understood for data subjects for the commons to be successful.

In the period which the interviews were conducted, the COVID-19 pandemic was taking place and so in consideration of the data protection and wider data-related challenges, analogies related to the pandemic were used. One interviewee explained how tensions exist when it comes to building a commons and considering data protection through the public or private sector, challenging existing norms when it comes to the use of our personal data: *"We need to understand that we are giving all this information to the private sector to run our lives or to help us run our lives. This used to be delegated to the public sector so let's think about it or at least discuss about it, and see which model we really want because when*

something happens, like coronavirus nowadays, no one is looking for answers in the private sector but looking in the public sector. So I think a lot of reflection needs to be done in this and a lot of dialogue with the citizens and a lot of speech needs to be there" (C2E2). In considering political and institutional barriers, one interviewee shared how a commons framework could be useful for opening up data and resources in a meaningful way: *"You have austerity destroying the public health infrastructure for a number of years. We have no valves, no ventilators, no masks, no protective equipment, and you see a mass of peer production groups that seek to solve these issues right? It is dialectic between the mainstream systems and increasing fault lines and then people self organising to find solutions beyond those bottlenecks basically"* (C3E1).

4.2.2 Overcoming data protection challenges

According to the experts, establishing relationships with data subjects and developing trust in both the commons framework and those who created the framework were important for the commons' success, particularly regarding personal data and data protection. While many of the commoners were engaged with their specific projects, transparency and clarity in the process of contributing to the commons can foster an environment for engagement to achieve a co-created commons outcome for individuals and groups.

One aspect is creating trust and establishing positive relationships between those who have an understanding of the data commons and data protection with those who do not: *"The main problem was trying to be careful in understanding each other in achieving the goals but it was a cultural problem when you interact with different people from different backgrounds, and that's a problem you have working with different people"* (C3E2). Another aspect is bringing the community together within the commons. One interviewee said: *"Two things were really striking, the first one is this binary process where either the user trusts you or doesn't trust you. But once they trust you, they give you everything. This is the direct consequence of, you know when you accept the terms and conditions of the services, that's the same way"* (C2E1). Another interviewee further explained: *"Other than the legal constraints [surrounding data protection and privacy, we didn't have any concerns that were raised]. This is one of the things that is really interesting and I think it is based on the trust. You have this social solidarity and there is this implicitly trust. If you break that trust, you are done"* (C6E1). This suggests that all stakeholders within the commons should be able to feel that they are being respected and treated as experts bringing in their own experiences, whether that

may be knowledge, perspective, or personal anecdotes regarding their data.

Regardless of the use case of the commons, it is important to understand community concerns, applied both to data protection and other issues. For data protection, this includes recognising the limitations of existing regulations and legal frameworks, such as the GDPR. One interviewee said: *"[Although, legally, you can ensure the process of deletion is followed,] you cannot tell people to forget something and they will forget. It was also something we realised with the GDPR law and our legal experts also discussed that"* (C2E1). These legal challenges regarding data protection also need to be considered throughout the commons development process, as interviewee C3E2 explained the role of their team was to *"deal with legal issues related to the goals of the project, fostering the making of the digital commons, including personal data"*. To overcome these challenges, input is needed from the community to assess the benefits and risks to the use of their personal data. However, Interviewee C5E1 explained that although the community was willing to engage, they felt unable to do so either because they didn't know how or because they had been approached in a manner which did not appeal to them. The type of involvement related to the sharing of citizens' data-related worries and the data protection issues they were currently facing to enable their data protection rights to be enforced. Another interviewee explained that the commons framework is useful for unpacking the socio-political challenges that impact the community, rather than specifically seeking a technological solution: *"We have just used the term [data commons] to introduce [stakeholders] to this kind of thinking to immediately hear them out and how they feel about the data society, about the smart city discourse et cetera and see within their context, in mobility projects in certain neighbourhoods, or energy transition, how they feel they want to deal differently with these technologies and how with urbanity or neighbourhood initiatives"* (C2E3). As a result, when overcoming data protection challenges within a commons, it is important to acknowledge the limitations of the law, technologies, and data. The commons could support different methods for allowing data subjects to choose their own personal data protection preferences.

4.2.3 Improving the commons

Addressing some of the general data stewardship challenges identified in Chapter 3.2, when discussing the usefulness and effectiveness of the commons, some interviewees expressed doubts. One said: *"I'm not entirely sure that [the project*

coordinators] actually achieved [their goals] in a reasonable sense because at some point there were too many challenges to resolve and we took some short cuts to reasonably put something forward for the demo, so there were lots of privacy issues that had to be solved later" (C1E1), emphasising the importance of timely development. Even in a commons, other stakeholders may be prioritised over data subjects, particularly when external financing and funding is involved: *"I often see the potential in people and areas in the project and then I have a hard line of what can and cannot be done and what the money was allocated for. So within our remit as an organisation moving forward, it will be a huge conversation with the much higher-ups than me about how do we deliver on our goals as set out in our original funding in a meaningful way . . . and I hope we have those conversations with people that are left out the most and working their way down to people who have access to things easily"* (C4E1). As a result, when establishing a successful commons, the scope of a specific commons is key to ensure that it is sustainable and balances the trade-offs between transparency and formalisation with more fluid and iterative ways of working: *"One of the other risks that came about was this transfer from a small project to a bigger project. These like growing pains are always difficult and the new definition of roles, the formalisation of rules, is really really interesting and also when it starts to make money. When there starts to be something to have, something to gain, something that people want then the interpersonal relationships really change and that can be a real risk in particular with group cohesion"* (C6E1). As part of the development process, if there is no community consideration, policy can be negatively impacted. From an interviewee, over 60% from a group of 50,000 people surveyed had never been consulted before: *"It is very concerning at a policy level where we are trying to make consulting decisions based on what the community want or what the stakeholders want or what the users want when the people we are hearing from are entirely unrepresentative of the local community"* (C5E1).

More specifically on developing a commons for data protection further as introduced in Chapter 3.4, to improve the commons for data protection, all interviewees suggested that collaboration across stakeholders and disciplines could overcome excluding data subjects and doubts about the effectiveness of the commons. This directly responds to the responsibilisation of the data protection process from Chapter 2.3.4. Working with stakeholders of different philosophical, technical, and social backgrounds helps everyone better understand how a commons could be implemented and be beneficial for reaching data protection goals: *"I think the literacy gap will be always there. You cannot rely on the public money going to literacy and to train people in terms of technology or whatever so the delegation of trust*

and transparency are the key" (C2E2). Another expert stressed the importance of inclusion: *"Low income and systemic inequality has left a lot of people not being able to access the internet like the rest of the world"* (C4E1). These considerations are also important when considering how data protection practices may be applied, on what mediums, and through what methods, particularly when addressing the reality on the ground one step at a time: *"The question for me is not how do we reach the end goal, the question is at the end of today, how are we one step closer to getting to the end goal at some point?"* (C6E1). Additionally, talking from the experience of COVID-19 and working online, some experts also believed that leaning into technologies and adopting digital tools in online-offline hybrid environments can make the commons more beneficial for a larger, more diverse group of people. One interviewee suggested: *"So it is one of those things that it is possible I think to develop social behaviour in digital means and then to develop commons in that way. . . . I think the reason why I have a different perspective on this is that almost all of my work is remote so I have super close relationships with people in the UK, working relationships with people I have met once live, but we meet online once a week and we chat through all the stuff we are working on. I mean we are all freelancers, this is our way of meeting at the water cooler"* (C5E1). Another interviewee emphasised the importance of building connections in physical environments as well as digital ones: *"How can we flagship different connections and [not just] the transition to digital, . . . but I am talking about organisations, and by organisations I sometimes mean people who live on the street and want to set up planters, maybe communities or neighbours are better"* (C4E1). As a result, improving the commons requires direct community involvement where the means for co-creation best reflect how they interact with the commons and their data, both through online and offline means as appropriate.

4.2.4 Building a commons for data protection

When considering the creation of a data protection-focused commons, there needs to be due consideration of the multidisciplinary nature of data protection and privacy. The commons could also recognise how the community's goals and how collective responsibility should be distributed.

One key point reiterated by many experts was the transition between theory and practice: *"Data commons is an idea that is hard to realise and our work was trying to make a tangible example of where this works and at least this had been tried and we'll see if it works or not"* (C2E1). The practicalities, given the strong relationship between

issues related to data protection and wider socio-political environments, must not be tackled in a silo or within one discipline only: *"People need this commons perspective because they are thinking about open data and balancing the protection of data so we should use the value of collecting data and findings but at the same time seeing to the sovereignty of citizens. It is one thing to understand what does this look like but in practice, how can we operationalise this?"* (C2E3).

According to the experts, action and collective responsibility were also key. An interviewee stressed the importance of action: *"[The commons] is a verb, it is commoning. It has the mindset of social solidarity and non-profit oriented. It is democratic and non-hierarchical"* (C6E1). Several interviewees mentioned that the purpose of a data commons needs to be clear as it is a choice. When building a data commons, more research needs to be done *"from legal, technical, social, political, economic areas of work"* and should include *"the vision of communities and people about what is at stake, what is this about, how it works, [and] how [data] has been managed"* (C3E2). Importantly, individuals and communities need to be encouraged and empowered to co-create: *"A lot of people do commoning but they don't know they are commoning. They don't have an identity that permits them to have, to exert directly power"* (C3E1).

Finally, it is important to consider what the goals of the community are and address those directly and collectively: *"I think that things will evolve because you know in free software, things evolved during decades from the beginning until now. ...I think something similar is happening with data now. We have the issue of tracking, contagion tracking apps, you know. This is creating a lot of debate about for example the data, why should the data be used for benefiting the community that is connected?"* (C3E2). Another interviewee explained: *"So it is related to those aspects, what do you give, what do you contribute to the commons and how will it be used? It is more on that issue that I have concerns. Who is deciding on how the commons is going to be used?"* (C1E1). Ultimately, as framed by one interviewee, the commons framework is seen as an alternative way of considering how a resource could be managed through transparently communicating risk and offering adequate protection within different hierarchical norms and rules as determined by the community itself: *"Whereas if you collectivise that risk, into an organisation like a union, then that body is about the same size as this other organisation and it is the same with commons. You are collectivising risk and you are also collectivising benefits because everything is distributed among the group in an equal way"* (C6E1). In this way, a data protection-focused data commons can support the community in making the most out of their data and personal data, without

disregarding the importance of protecting that data in the first place.

4.2.5 Interviews summary

Our interviews indicated that data protection within existing commons frameworks was predominantly considered only in terms of control and sovereignty over data subjects' personal data. Although the decision to use a commons was to provide certain levels of control and transparency over how data were collected, used, and processed, there were limited applications to wider data protection principles such as those relating to informing data subjects about their rights and the ability to exercise those rights against data controllers, echoing our illustration of current data stewardship solutions from Chapter 3.2. According to the experts, establishing strong community relationships to develop trust in both the commons framework itself and those who created the framework was important for the commons' success. While many of the commoners were fully engaged with their specific projects, transparency and clarity in the process of contributing to the commons can foster an environment for engagement to achieve a co-created commons outcome for both individuals and groups. Openly acknowledging the legal and technological limitations within data protection, such as those outlined in Chapter 2.3.1, can also result in a more supportive and collaborative environment for creating data protection solutions. Interviewees also expressed their doubts over the use of the commons framework for their specific projects as certain assumptions were made about its ability to be put into practice. However, all interviewees suggested that collaboration through online and offline means across disciplines could overcome this challenge. As expressed in Chapter 3.4, the commons can support data protection and allow data subjects to extract value from the knowledge generated from their data without sacrificing privacy where the limitations of legal protections or technological innovation are communicated. Working with stakeholders of different philosophical, technical, and social backgrounds helps everyone understand how a commons should be implemented and could be beneficial for reaching data protection goals.

Based on our interviews, we find that the polycentric governance of the commons and ability to center discussions around data protection within the commons demonstrates that a data protection-focused data commons can be appropriate as a socio-technical framework for data stewardship as long as the purpose of the commons is clear to data subjects. Crucially, reflecting concerns identified in

Chapter 2.4, socio-technical requirements as part of wider data management and stewardship considerations in a commons can be implemented to encourage the co-creation of data protection solutions for the common good.

4.3 Adapting the IAD Framework for Data Protection

In considering creating a data protection-focused data commons, the experts identified important considerations throughout the development process. While their experiences highlight the importance of including data subjects early on in creating a commons, there remain open questions about how to implement and develop a data protection-focused data commons. To ensure that the data protection-focused data commons is appropriate as a socio-technical framework for data stewardship, we apply the interview themes and questions considered by the experts from their interviews identified from Chapter 4.2 to the IAD framework as discussed in Chapter 2. The application of our interview findings to the IAD framework allows us to implement the inclusion of data subjects in the commons development process, addressing the data stewardship challenges identified from our analysis in Chapter 3.2.1.

In creating a socio-technical commons for data protection, we found that collaboration across stakeholders and disciplines could overcome excluding data subjects and doubts about the effectiveness of the commons. Building the commons includes recognising the theoretical, philosophical, and psychological aspects of people's relationship with privacy. Those in the commons could then be allowed to express their views, where these perspectives are drawn out from the "Background" section in the IAD framework. These perspectives have also been included as part of the commons' goal based on its use case. When establishing a commons, the purpose of the commons needs to be clear, as identified from the "Goals and Objectives" section, because the use of the commons model is a choice, and that clarity allows for new iterations of the commons to best suit data subject needs. The commons could include the vision of communities and people about what is at stake, what it is about, how it works, and how data have been managed, as included in "Commons Community Members". Ultimately, commoning was identified as a verb, where the community has to actively participate in the development process and its application, and is necessary for successful co-

creation and participation. As the commons has been identified as a means of community collaboration, opportunities for feedback and iteration such as chats, forums, and public consultation processes need to be included in the commons process. Importantly, data protection within the commons itself should also be transparent and reflect the needs of data subjects, as reflected in the "Outcomes" section. The data protection IAD framework is as follows:

Background

- Identify the background context of the commons.
  - What are the relevant potential data- or data protection-related events, such as data breaches or privacy scandals that may have prompted the interest of creating a data protection-focused data commons?
  - Are there any other background considerations to include such as the introduction of new data and data protection regulations, the uptake of certain technologies, or the introduction of sector-specific changes in data or data protection?
- Identify the existing norms, standards, and guidelines when it comes to data and data protection considerations in this background context.
  - What sector-specific guidance such as code of ethics or code of conduct documents, information provided by data archivers or data protection officers, and data sharing documentation are there for this use case?
  - Are there any forms of data licensing that may be most common? Are there open data principles within the commons use case?
  - What are the limitations of some of these norms? Are there examples or areas where the community could have more support?
  - Have there been difficulties in balancing openness and protection of data in this background context in the past?

Data Attributes

- The data and personal data that is part of the commons.
  - What is the data and metadata to be protected by the commons?
  - Is the data special category personal data as defined by the GDPR?

  - How is the data created and collected?
  - What is the data and digital infrastructure (internal or third-party) used to store this data?
  - How private or public is this data? Who is it being shared with and do users have control over this?
  - Who is involved with the process of creating, collecting, and storing this data?
  - Are there specific technologies or tools involved in this process?
  - How much control do the creators of the data have in the current background context?
  - What are the risks of sharing personal data given this information and how could such data sharing generate benefits for data subjects themselves? Are there opportunities for data subjects to choose whether they are willing to take that risk?

Commons Community Members

- Who are the community and commons members? What are their roles?
- What is the degree and nature of openness with respect to each type of community member and the general public?
- Who are the other stakeholders that may be involved with the data and data protection process of the commons?
- What are the relationships between the different stakeholders? Are there potential issues of trust between them?

Goals and Objectives

- What are the goals and objectives of the commons and its members, including obstacles or dilemmas to be overcome?
- What is the history and narrative of the commons?

Managing and Governing the Commons

- Identify the action arenas and the goals and objectives of the commons.
  - What scenarios are the commons aimed at dealing with, such as specific use cases of personal data, as well as linking them to previously identified goals of the commons?
  - How do the community and commons members relate to other stakeholders? Is data shared between them used similarly or is the data used for different purposes?
  - Are there existing data protection mechanisms, such as exercising data subject rights, data protection impact assessments, or introducing DPbD, that are applied to achieve the goals and objectives of the commons?
  - Which data subject rights can the community and commons members exercise in relation to their personal data?
  - Have principles of purpose limitation when collecting and using personal data been respected?
- Determine the governance mechanisms of the commons.
  - Are there membership rules to being part of the commons? Do participants have to be concerned about a specific means of using data or have certain questions in mind which they hope the commons can help solve?
  - Do commons participants have to contribute certain forms of knowledge or data to be part of the commons?
  - What are the means to resolve conflicts within the commons when it comes to finding solutions?
  - Are there mechanisms for sanctioning rule violators?
  - Are there technologies or digital infrastructures that could help with mechanisms of governance?
- Identifying decision-makers and experts.
  - Beyond the community and the commons, are there other decision-makers and stakeholders that act as experts within the commons? How are they selected?
- Institutions and technological infrastructures that structure and govern decision-making.

  - Are there places or platforms where the commons could take place, physical or digital?
  - Are these institutions and infrastructures internal or external? Is a specific infrastructure being developed for this commons?

- Establishing formal or informal norms that govern the commons.
  - Does the commons have guidelines that help establish the code of conduct or provide a red line on what is unacceptable within the commons community?
  - How can the commons community help create and develop these guidelines?
  - Are there existing examples of norms that can be applied to this use case?

Outcomes

- Benefits of the commons.
  - What are the benefits of participating in the commons when it comes to protecting personal data?
  - What should the community expect when they participate in the commons?
- Costs and risks of the commons.
  - What are the risks of participating in the commons when it comes to protecting personal data? Are there risks of further data breaches or privacy problems?
  - Are there risks specifically related to introducing technologies into the commons?
  - Are there mechanisms in place to protect against violations of regulations such as the GDPR from within the community?

4.4 Summary

In this chapter, we have reviewed the feasibility of a data protection-focused data commons and assessed how a commons can be created in practice. Our study has

indicated that a data protection-focused data commons is appropriate as a socio-technical framework for data stewardship, where the commons can help support data subject collaboration and the redistribution of power between themselves and data controllers. We note the following:

- From our interviews with commons experts, we identified the data protection challenges for creating a commons, how to overcome them, how to improve the commons more broadly, and the important requirements for building a data protection-focused data commons.
- We find that working with stakeholders of different backgrounds and perspectives can support a commons' implementation.
- To support the deployment of a data protection-focused data commons, we propose requirements by adapting the IAD framework based on the themes identified from our interviews and addressing concerns raised from Chapter 3.2.1.

Having shown that existing commons methods can be improved and applied to the creation of a data protection-focused data commons, in the next chapter, we discuss how a data protection-focused data commons could be practically implemented in policy measures and set out the requirements for implementing a data protection-focused data commons.

CHAPTER FIVE

ESTABLISHING A DATA PROTECTION-FOCUSED DATA COMMONS

In Chapter 2.3, we identified the data protection challenges that have emerged after the GDPR came into force, while in Chapter 3.2, we illustrated how data stewardship frameworks that aim to protect data subjects' personal data may leave them out of the data protection process and inadequately support the mitigation of potential data-related harms. In response to these challenges and addressing the first research question identified in this thesis, our interviews in Chapter 4 found that the commons can be adapted with a focus on data protection to encourage data subject engagement and the co-creation of data protection solutions based on data subject preferences.

In this chapter, we build upon our interview findings from Chapter 4.2 and embed these into the process of establishing a data protection-focused data commons. These themes include incorporating multidisciplinary considerations as part of the commons process, considering socio-technical requirements as part of wider data management and stewardship, and establishing strong community relationships to develop trust. By embedding polycentricity into the method of creating a commons for data protection, we address the multidisciplinary challenges to responsibilisation identified as through the "problem of many hands" and "cooperative responsibility", outlined in Chapter 3.1. This *ex ante* process further introduces critical ethical and interdisciplinary considerations

for data protection that go beyond the IAD's analytical *ex post* assessment of the commons (Chapter 3.3). We reinforce this through directly responding to the concern that data-related rights may not be fully realised within existing data stewardship frameworks, as identified in Chapter 3.2.1, by incorporating data rights considerations into the creation of a data protection-focused data commons.

To establish a data protection-focused data commons and differentiate it from existing commons as identified in Chapter 3.3, it is important to first define what a data protection-focused data commons is and how its aims for supporting the co-creation of data protection solutions for the benefit of data subjects, the second thesis research question, can be achieved. To do this, we first apply the theory and principles of the commons with the themes identified from our interviews to outline how the data protection-focused data commons can encourage data subjects to co-create and collaborate on data protection solutions. Next, while the application of data protection into a data commons is possible, to ensure that it can be implemented in practice, we adapt the data protection IAD framework created in Chapter 4.3 into a policy scaffolding to ensure that the socio-technical aspects of data protection are included, as previously identified in Chapter 3.4. The policy scaffolding will serve as a practical measure that embeds data protection rights and principles to the commons framework that can be understood without extensive knowledge of the commons. Finally, we demonstrate how the data protection-focused data commons can include and support data subjects in protecting their individual and collective data by applying the commons to use cases.

In this chapter, we make the following contributions:

- We outline how a data protection-focused data commons could be mapped from commons principles.
- We set out the requirements for implementing a data protection-focused data commons.
- We establish a policy scaffolding for establishing a data protection-focused data commons to support the practical application of socio-technical considerations in data protection.
- We suggest scenarios in which a data protection-focused data commons could be applied to support data subject engagement and the co-creation of data protection solutions.

5.1 Defining the data protection-focused data commons

Using the theory and principles of the commons illustrated in Chapter 3.3, we suggest that a commons for data protection can be created to allow individual data subjects as stakeholders to collectively curate, inform, and protect each other through data sharing and the collective exercise of data protection rights.

Before setting out the use cases in which a data protection-focused data commons can be applied, we define and clarify why the framework is important in context of privacy and data protection, which we introduced in Chapter 3.3. In a data protection-focused data commons, the common property is personal data from data subjects that is used for a specific purpose as well as any relevant information related to how data is collected, used, and managed within a use case, where the commons incorporates existing legal and technological structures as well as data subject input and preferences. The data protection-focused data commons aims to help contextualise privacy beyond control and move towards privacy as ability and as a state, enabling a mechanistic expectation that addressing differences will make more people comfortable through relationships of respect [210] and the necessity for collective rights [197] both before and after data is collected. A commons encourages iterations of individual and group data protection objectives that can be different, personalised, and change over time [146]. Figure 5.1 shows how data subjects are the focal point of the commons, while other stakeholders are bound by the data subject's desire for the protection of their personal data according to data subject preferences. Within the framework, the data subject can also interact with data controllers, data managers, researchers, and civil society for generating their data protection outcomes. A commons developed using Ostrom's design principles, as described in Chapter 3, is useful because of the vast number of stakeholders that have a diverse set of opinions, problems, and preferences on how the data subjects' personal data is managed as identified in Chapter 2.3.2. Polycentricity is also recognised, where the complexities of different stakeholder tensions are transparent and considered within the commons itself.

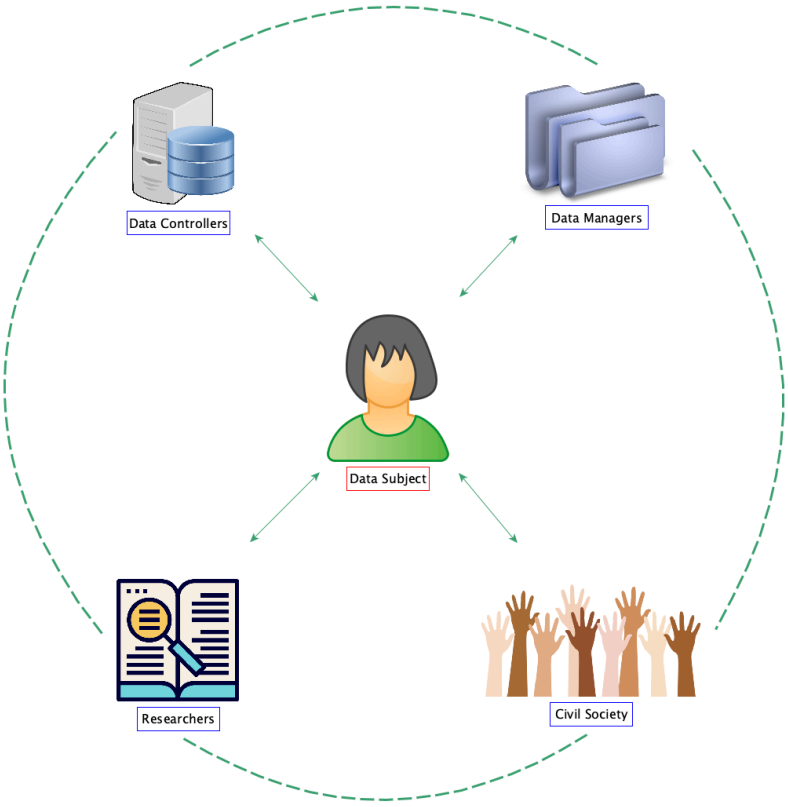


Figure 5.1: In a data protection-focused data commons (green), the data subject is at the centre. In this framework, the data subject and their personal data is the main priority of the commons, and other stakeholders are only considered in context of the data subject's data protection. The different stakeholders represent the polycentricity of all the systems which have influence over data subjects.

5.2 Data protection-focused data commons requirements and stakeholders

To implement a data commons, various legal and technological components as noted in Chapter 2.2 need to be created for stakeholders to be engaged. Figure 5.2 shows how a data subject specifies to what extent they would like their data to be protected based on existing conflicts and challenges pre-identified within the commons for the use case. In addition to data subject preferences, using information such as data controller policies, research papers, and input from civil society, a data subject specification is created and used to inform their data protection outcome that is generated from the system. For example, the data subject-centred outcome could ensure that the data subjects' data is only to be reused by data controllers and other stakeholders that the data subject has

expressed in their preferences and not by specific third parties. These stakeholders would automatically be notified of the data subject's preferences. This allows the data subject to set their own limitations of how their data is used as opposed to it being decided by the platform's requirements or defaults. As the outcome is data subject-centred, decisions ensuring the protection of the data subject's personal data may override existing preferences, policies, or standards set by other stakeholders. Data subjects can return to and review their outcome, add their data subject experiences to the commons, and participate in the co-creation process at any time. Data controllers can address new risks before collecting data, minimise the potential for data breaches, and meet stakeholder demands. Other stakeholders, such as researchers and civil society, can participate in the commons to make the data sharing process more transparent, support exercising group rights, and provide information and standards such as FAIRsharing [205]. These requirements aim to decrease the power imbalance between data subjects and data controllers. Mapping out the development of a commons into Ostrom's CPR design principles, a commons will be clearly defined based on its use case, where each stakeholders' role is detailed. All data subjects that would like to find out more information about the commons use case, contribute, or co-create are free to participate in the commons. Any bad practices, unethical behaviour, and data breaches will be identified by the commons system, with the remedies updated as stakeholders respond. Data subjects and other stakeholders can collaborate and establish their own norms, such as co-creating data sharing practices which promote DPbD and facilitate data reuse for data research projects amongst a group of researchers.

5.3 Data protection-focused data commons scaffolding

Following the adaptation of the IAD framework for creating a data protection-focused data commons in Chapter 4.3, we establish a process, described as a policy scaffolding, for commons policy-makers based on the themes outlined by the experts. This process is described metaphorically as "scaffolding" as it suggests providing support for creating a new or iteratively repairing a structure, in this case, a data protection-focused data commons. Scaffolding is also used to access places that are difficult to reach, as represented by the challenges of creating a commons outlined by the findings from our interviews where difficult questions

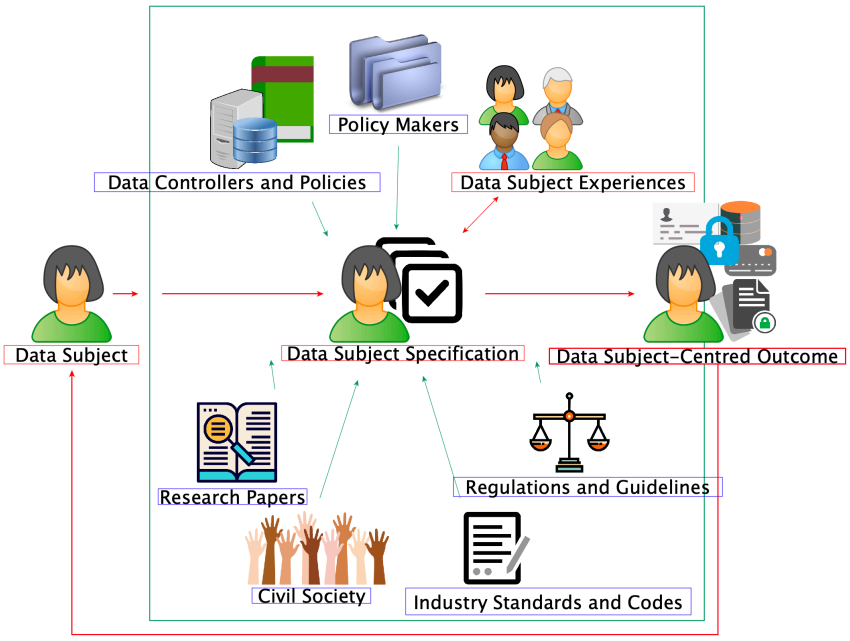


Figure 5.2: In a data protection-focused data commons (green), the data subject specifies to what extent they would like their data to be protected based on existing conflicts and challenges pre-identified within the data commons for the use case (red). No prior knowledge of existing law, norms, or policies are required. Along with stakeholder information (blue), the data subject specification is then used to inform their data protection outcome that is generated from the system. As the outcome is data subject-centred, decisions ensuring the protection of the data subject's personal data may override existing preferences, policies, or standards set by other stakeholders. Data subjects can return to and review their outcome, add their data subject experiences to the data commons, and participate in the co-creation process at any time.

first need to be answered to justify for what and how a data protection-focused data commons is to be used. The scaffolding metaphor has been used for problem solving as part of education [269] as well as to understand social interactions and engagement through inquiry [203]. As a tool, scaffolding has been considered in open-ended technology based environments [135]. The data protection-focused data commons policy scaffolding is intended to act as a starting point for creating a data commons for data protection, where data subject engagement is embedded as a necessary component of the commons development process.

Building upon the findings from the interviews, we adapt the themes and solutions identified to the IAD framework from Chapter 3.3.1 and create a policy scaffolding to outline a process that supports commons policy-makers in developing a data protection-focused data commons. The adapted IAD framework from the previous

chapter, aimed at supporting commons practitioners and those developing the commons, can be simplified and translated to focus on meeting specific policy agendas through direct data subject engagement and critical assessment of the data protection concerns of data subjects. For example, the IAD framework focuses on identifying the data attributes to map out the data infrastructure considered within the commons. In contrast, the policy scaffolding focuses more on the scope and impact that the commons intends to have to clarify how the commons' aim could fit into wider policy agendas in context of data subjects. This can simplify the practical application and adoption of the commons for data subjects and other commons stakeholders without the burden of understanding the theories and principles behind the commons. The policy scaffolding transforms concepts within the IAD framework into an actionable, chronological process for planning, creation, and maintenance of a commons. The scaffolding also integrates key questions and themes identified by commons experts from Chapter 4, such as how data protection challenges can be overcome and how data subjects can be encouraged to participate in the commons process, to put polycentricity into action.

By using a data protection-focused data commons scaffolding, the key aspects of implementing a commons can be identified and more easily put into practice by policy-makers and applied to specific use cases for local communities. Establishing policy documentation and creating a scaffolding for policy-makers encourages data protection to be more holistically considered for each use case to allow for co-creation, engagement, and participation for all stakeholders within a commons. As the process of creating a commons requires having data subject participation from the beginning, this scaffolding places their considerations throughout the commons development process by directly asking policy-makers questions that can only be answered by data subjects and so require that they be included in the process. Taking the analysis questions developed by McGinnis (outlined in Chapter 3.3.1) and incorporating them into this scaffolding, this set of analysis questions is more suitable for creating a policy scaffolding as it provides a more holistic view of the commons development process beyond the IAD framework while also including direct data subject engagement. Given that the IAD framework and policy scaffolding are not legislative proposals, the framework may be considered a means to initiate and use the commons within existing legal and socio-technical infrastructures to encourage engagement and participation. We also address the commons policy implications noted by Sanfilippo et al. and elaborate on how policy-makers can promote appropriate

information flows while protecting personal data [204].

The scaffolding is split into four parts, mirroring the interviews analysis in Chapter 4.2: identifying the commons use case and the data subjects, scoping and information gathering for developing the commons, building the commons, and sustaining the commons. Each part reflects the themes drawn out from our interviews, particularly on how a commons can be improved for data protection. The scaffolding also incorporates advice from experts on how to meaningfully involve data subjects at each stage of development, ensuring that the aim of the commons supports data subjects in creating data protection solutions. The scaffolding is as follows:

1. Identifying the commons use case and the data subjects

The first step for policy-makers when creating a data protection-focused data commons is to identify the use case to which a commons can be applied and the data subjects that the commons aims to benefit. When examining a data commons use case, a data protection-focused data commons could serve as a new public consultation mechanism for policy-makers and help identify data protection best practices to incorporate into policy. Directly incorporating data commons policies into consultation work allows data use, sharing, and methods for protection to be transparent, ensuring that data subject perspectives are considered in the process:

- What is the data protection issue for the data subjects for the use case?
- Is the issue one that relates more to finding a policy solution or does it require a wider scope in identifying an underlying problem? What could have caused this issue and was there any event that may have exemplified it?
- What resources are already available to support data subjects and how should this information be presented to them within a commons to make it more accessible?
- Which data subjects would be invited to participate and engage in the public consultation process?
- How should data subjects be included?
- What value does better data protection for this use case bring to the data subjects involved and also to the data itself?

2. Scoping and information gathering for developing the commons

The next step involves considering stakeholders and including data subjects early on in the process. For data subjects, when creating and using a data commons, writing new community policies as well as using existing data protection policies, such as regulations and institutional policies or codes, can support them in co-creating data protection responsibilities for and alongside other stakeholders. Guidance could also be provided for data subjects if they wish to co-create policies within the data commons:

- Who are the other stakeholders with more power over personal data compared to the data subjects?
- Who are the other stakeholders that can support data subjects and provide more information for them?
- What relationships do these stakeholders have between each other? Are there specific stakeholders that are dominating what happens within the commons use case identified?
- How would information, advice, and participation from different stakeholders, including data subjects, be included during the commons development process?
- What are the wider data protection and privacy issues (social, technological, and philosophical) that relate to this use case?
- What solutions, if there are any, have data subjects tried in an attempt to solve the data protection issue identified?
- Are there similar commons or data stewardship examples that can help support the creation of this new commons?
- What laws, technologies, or policies have been developed for this use case that could support better data protection practices? Are these enforced under the law, codes of conduct, or the community? Are these effective?

3. Building the commons

Once the stakeholders have been identified and a preliminary blueprint has been drafted, the next step involves creating and building the commons by addressing

some of the issues previously identified in more depth. This stage involves more involvement to ensure that the commons development process can be iterative and best reflect data subject preferences. Some of these questions include:

- As part of the commons process, how will you find out what data subjects' data-related worries are and support their data protection rights?
- How can the wider data protection and privacy issues identified previously be addressed either as part of the commons itself or from the commons development process?
- Within the commons, what mechanisms can help develop trust between stakeholders, particularly for data subjects?
- What assumptions related to the use case need to be addressed and corrected? What baseline information should data subjects have to best help them co-create the most suitable data protection solution?

4. Sustaining the commons

Finally, policy-makers could consider how the commons can operate in the long-term with other stakeholders as well. For example, it may be useful to include data controllers so that they can more clearly understand what data protection requirements are preferred by data subjects. For DPAs, policies established around creating a data commons for specific use cases help ease their burden of enforcement through preventative data protection measures, *ex ante*, before data is collected as opposed to remedying data breaches, *ex post*. Additionally, establishing a data protection-focused data commons framework in policy encourages policy-makers to reconsider current balances of power between data subjects, data controllers, and other data protection stakeholders, taking into consideration the data ecosystem in the long-term for socio-economic benefit by increasing the value of data:

- How can the commons be sustainable in the long run? What can be done at the development and implementation stage to ensure that the data protection issue can be better managed and solved?
- Are there particular platforms or infrastructures that can help host the commons and ensure that it is as accessible as possible?

- How can collective responsibility be demonstrated within the commons and how can its reach be maximised?
- What other stakeholders can be brought in to help support and sustain the commons?
- How will you know when the data protection goal of the commons is achieved? How can this be measured?

In addition to the practical application of the data protection-focused data commons, the scaffolding differentiates a commons centred around data protection with existing data commons in that legal and socio-technical considerations of data protection are fully considered from the perspective of data subjects in the former. While the data protection-focused data commons redistributes the value of data back to data subjects similar to other data commons, it also supports iterative development through consultation processes with data subjects to ensure that any data protection solutions are co-created and collaborative in nature. The scaffolding also introduces the consideration of measurable outcomes and objectives that data subjects can identify for themselves as individuals as well as collective goals that are co-defined by all stakeholders within a commons. The combined use of the IAD framework illustrated in Chapter 4.3 and the policy scaffolding in creating a data protection-focused data commons enables the consideration of practical socio-technical requirements and longevity of the commons as applicable within wider community and societal aims respectively.

5.4 Use cases

To demonstrate how a data protection-focused data commons could increase data subject engagement with the data protection process through collaborative and co-created means as noted in Chapter 3.4, we apply the commons to the use cases of data archiving and online learning based on both the IAD framework from Chapter 4.3 and the policy scaffolding outlined in the previous section.

5.4.1 Data archiving commons

We now outline how public data archives can benefit by applying the use case to a data protection-focused data commons. We assess how a commons could address stakeholder issues by increasing accountability for data subjects' personal data

and general archive data, encourage collaborative data protection solutions, and allow for data protection to be an iterative process.

Given the digital nature of resistance movements and networked protests such as the Arab Spring and Occupy Wall Street [242], for activists as data subjects, participating in a data protection-focused data commons allows them to understand social media platform archival policies, identify an appropriate platform that best balances their data protection concerns with the need to disseminate information, and limit the extent to which they wish to share their data for research or third-party purposes. As outlined in Chapter 3.3.2, socio-political sensitivities associated with this data can have personal and group ramifications to the data subject. For example, the data subject's data or posts may no longer be in the public domain, but have been curated for research archives inaccessible to the data subjects themselves, resulting in politically sensitive information remaining visible or searchable without their knowledge or consent. The data protection-focused data commons can therefore be useful for supporting activists' data protection rights while also enabling the co-creation of data protection solutions for collective benefit.

For a data archive data commons, we apply the Umbrella Movement, Hong Kong's 2014 democracy protests for universal suffrage, to the theoretical commons framework to demonstrate our use case as existing research has demonstrated the data protection and data research challenges for this social media archive [241]. This use case is applicable for the commons as the iterative nature of the framework is able to reflect temporal changes that could impact data subjects and how they would like to protect their personal data, such as the rapid socio-political and legal developments in Hong Kong between 2014 and the time of writing in 2022 [120] [145] [19] [153] [156]. This reflects the commons' ability to help tackle issues related to data protection and wider socio-political environments as identified from our interviews in Chapter 4.2.4. Created before the collection of data, the commons allows activists and potential research participants as data subjects to see what and how their data will be shared with data controllers and researchers, raising and addressing any concerns respectively. Applying Ostrom's design principles, the commons for data archiving will have clearly defined boundaries as to the kinds of data and metadata it will archive and when such archival will cease. In this example, data subjects would like to publicise their experiences from the Umbrella Movement on a platform in the

public domain. To participate, the data subject can identify the most applicable commons for their purpose by searching for keywords such as social media, data archive research data, and data reuse. The identified commons would include information about data controller policies on personal data, research and archiving, other data subjects' experiences and outcomes from exercising their data subject and information rights, recent news and scandals on data controllers, and expert and researcher findings from their work based on relevant topics and tags. This allows the data subject to identify what settings there are for preferences such as limiting the audience, how information can be published in public and in private, whether data can be deleted, how published information could be used (by who, how, and the process), what the intellectual property policies are for the published information, and how other data subjects felt about the platform's responses to information rights based on their experiences. This responds to the question "What resources are already available to support data subjects and how should this information be presented to them within a commons to make it more accessible?" as identified in the first step of our data protection-focused data commons scaffolding in the previous section. Without a data commons, the data subject would have to search for this information independently, looking for forums for information.

With a data protection-focused data commons for data archiving, existing standards and review mechanisms such as research ethics board reviews, funding body requirements, and institutional policies can be integrated into the commons, acting as the first level of safeguards for data protection for data subjects in the future. This reflects the value that better data protection for this use case can bring to both the data subjects and to the data, as noted in our data protection-focused data commons scaffolding. Researchers that work in data archiving and examine data protection practices on archive data reuse can offer privacy principles for individuals and organisations to adhere to. Although ethics approval for researchers to gain access to the archive may be granted by institutions, if the research data reuse by third parties is also granted, the participant as the data subject may not know what stakeholders have access to their personal data and for what purpose. In cases where ethics approval is dubious, such as where data is shared without authorisation or full anonymisation, researchers may take their work outside of institutions and use data subjects' personal data in commercial ways, as was the case of Cambridge Analytica [31]. This can be mitigated with a commons where future researchers looking to use the data archive can utilise it to

see what data limitations have been discussed and set by data subjects, proceeding with reuse if those requirements are met. The inclusion of wider data protection and privacy issues as they relate to the social, technological, and philosophical relationships data subjects have with personal data supports the scoping and information gathering section from our scaffolding in the previous section.

After the data subject uses the commons to identify the most relevant framework for their purpose, Twitter for example, based on different stakeholder knowledge, the data commons uses the information and prompts the data subject to select a few preferences by answering questions based on the conflicts and challenges that have arisen from them. These questions could include "Do you want your data or posts to be publicly archived?", "Is this archive available to researchers?", "If your data or posts were deleted at a later date, would you want to notify researchers of such request, for example to not include your data in future studies?", and "Do you want a mechanism to hide all or some of your data and posts?". Based on the data subject's responses to those questions, the data commons chooses platform settings and data actions for the data subject that best aligns with their aims. These may override platform policies based on data subject requests. For example, during the Umbrella Movement, Twitter decided to change its historical archive policy regarding the removal of deleted tweets and made all public tweets findable [90]. This information would be automatically reflected in the commons through technical means, notifying the stakeholders in the system. Although Twitter automatically removes deleted Tweets from its data archive, if a data subject would like for that information to be kept in certain pieces of work, researchers have a right to retain such data until further notice by the data subject. Requests by data subjects could be specific, such as any Tweets that include the term "Umbrella Movement" can be kept while ones with "universal suffrage" or "Hong Kong independence" can be removed. Researchers would be notified of these preferences. The system can then be updated to match personal preferences with secondary resources to create a more comprehensive picture of what preferences data subjects would collectively like with regards to data archiving, sharing, and reuse.

Collective participation in the commons with other independent stakeholders can also support data subjects in co-creating data protection solutions. For example, experts such as Tromble and Stockmann who researched on this topic can advise data subjects on what they can do in light of new policy changes as well as data

controllers on how to address any concerns raised. Even without misuse and with GDPR Article 17 RtbF, certain forms of data archival can make removing personal data difficult, particularly if such data has already been reused in research. If Twitter suffered from a data breach and released the personal data of data subjects tweeting about the Umbrella Movement, in a data commons, the breach can be addressed by supporting data subjects in exercising the RtbF and sending notifications to data controllers to request their data be removed. Researchers who have used the affected datasets and data would also be prompted of the breach. They would then be required to issue corrections in their work and remove identifiable data in relation to the data subject from their data archives should the RtbF be exercised. Not deleting the relevant data could result in terminating access to the commons and future data. Automatic detection of subsequent attacks caused by the data breach can also prompt the system to alert and support data subjects to exercise their data subjects rights if they haven't already, as well as look for new alternative platforms that support data protection practices according to the data subject's preference.

In deciding the best platform and settings for the data subject's purpose of broadcasting the Umbrella Movement, responding to questions identified in the "Sustaining the commons" section of our data protection-focused data commons scaffolding, further advice is also provided on how data from the data subject can be best protected. This includes: setting up an account with a disposable email, having an anonymous platform username, setting up tools that can automatically delete the data subject's posts, links to how to exercise information rights on the platform, and the successes and failures of other data subjects in this regard. This information is saved in the data commons and is accessible by the data subject at any time. Any information that the data subject has gathered can also be put into the data commons. Collective participation by different stakeholders further allows research participants to assess these forms of use and curate the data archive data for themselves, engaging with their own research interests while participating in the community interest. Using the IAD framework for data protection as a reflective instrument, the data commons acts as a new means for data management where the reasons for use, limitations on reuse, and recourse after data is aggregated and anonymised are all contained within one ecosystem.

By creating a data archiving data protection-focused data commons, established commons methodologies and organisational structures are built into the data

protection-focused framework while enabling individuals and groups whose data forms part of the datasets to determine how their personal data is used. Given the significant socio-political risks that data subjects may face should identifiable data be shared without their knowledge, a commons could support their agency over their personal data. Using the data protection IAD framework and policy scaffolding, we can specifically identify how the commons as applied to data archiving can enable collaborative data protection solutions, where the polycentric system helps increase awareness of different stakeholder perspectives. This includes data protection law and policy, data subjects and their rights, data controllers, data managers, archivers, and researchers, developing a more comprehensive understanding of how personal data can be protected in archival settings. The commons simplifies the data protection rights procedure by including information, instructions, and templates on how rights could be collectively exercised, giving data subjects an opportunity to engage with and shape data protection practices that govern how their archived personal data is protected, particularly as the public nature of the data archive may result in additional potential harms.

5.4.2 Online learning commons

Online learning and tutorial recordings are another example that reveal data-related power imbalances between data subjects and other stakeholders. Given the shift towards online teaching and remote learning as a result of the COVID-19 pandemic as identified from our case study in Chapter 2.4.3, for students as data subjects, participating in a data protection-focused data commons allows them to understand their school's or university's policy and external organisations' guidance when it comes to collecting, processing, and sharing their personal data related to online learning. Students can pose questions to experts, raise any questions about data protection to staff, review their consent decisions on whether to agree to tutorial recordings, and exercise their data protection rights should they wish to do so. While staff can also be considered as data subjects, our use case focuses on students due to the unique stakeholder tensions and numerous forms of educational, academic, and administrative data students may have in relation to staff, HEIs, and future employers, making it more difficult for them to object against the collection of their online learning data. Using the data protection-focused data commons scaffolding, an online learning data protection-focused data commons can be created to support the commons process in allowing for

collaboration and engagement for creating data protection solutions.

Given that online learning data is collected, managed, and shared by many different stakeholders within HEIs [170], the data protection-focused data commons can support data governance for online learning by collating all forms of data-related information provided by the university and independent external guidance for online learning, and equip students with the information to exercise their data protection rights if they wish to do so. The commons, without data protection considerations, has previously been theoretically adapted to the university environment. Madison illustrates that as universities continue to evolve, the nature of the university may change from a knowledge to a data-oriented institution, resulting in the conflation of data as knowledge [143]. As a result, wider change in institutional governance may also be required. For HEIs to manage their resources for maximum benefit and minimal social and private harm, they could consider the knowledge commons as outlined in Chapter 3.3.1. HEIs could examine data governance beyond intellectual property rights and be open to multi-stakeholder engagement when creating university policies and meeting third-party obligations for academic data. Although the risks of data collection, sharing, and security are not explored, Madison offers insights into how university data could be managed as a commons via strategies of openness, sharing, and polycentricity, but with contextually-appropriate elements of intellectual property management and data exclusivity. This assessment of the types of data collected and data management process as part of online learning reflects the "Data Attributes" considered in the data protection IAD framework. As a result, to increase student agency in protecting their personal data, a commons could be created to support collaborative means for them to meet their data protection preferences with the knowledge of their institutions' data protection practices and data protection rights. By extending this adaptation of the university environment to a data protection-focused data commons, an online learning commons can directly provide students with the related university policies or best practices, give students the ability to decide whether they consent to certain collection and processing of data, and support students to exercise their rights to the university's DPO.

An example use case application for an online learning commons could be a data protection-focused data commons applied to support whether students want to opt-in to tutorial recordings and determine how their recordings can be used or reused, identifying the commons use case and data subjects following

our scaffolding. Given the prevalence of tutorial recordings and the use of e-proctoring services identified in Chapter 2.4.3, by applying the data protection IAD framework and policy scaffolding to an online learning commons, greater trust and transparency on how student data is used can help support them as data subjects. Responding to the "Scoping and information gathering" section of the scaffolding, this allows them to learn about and exercise the data protection rights that they have as both individuals and groups. Particularly where there are power imbalances between them, staff, and their university, the commons can identify the tensions between different stakeholders and address them following university policies, understanding the privacy preferences available within the online learning platforms deployed, and provide instructions on how to seek data protection recourse. To participate in the commons, the data subject can access information about their HEI's data protection policies to see how their data is collected for online learning. Other information could include external documentation on online learning best practices, such as those identified in Chapter 2.4.3, different tutorial recording platforms, and the DPO contact, as assessed under the "Commons Community Members" section of the data protection IAD framework. This allows the data subject to see what available options they have when it comes to being recorded and whether any recorded data could be deleted. To support data subjects who are unsure what to decide for their data protection preference, answering a quiz with questions such as "Do you plan to reveal sensitive personal data?", "Will you re-watch the tutorial recording?", and "Are there topics you would avoid discussing if the tutorial is recorded?" could automatically recommend to the data subject whether they should or should not consent. The data subject is also able to change the suggestion at any point. To remove peer pressure from other data subjects when considering whether an individual should consent, the consent vote could be made anonymous and take place before the tutorial. This supports our "Building the commons" scaffolding section, where questions related to understanding data subjects' data-related worries and supporting a baseline understanding of data in context are considered.

An online learning data protection-focused data commons also simplifies the data protection rights procedure by including information, instructions, and templates on how rights could be collectively exercised within the online learning context. This gives data subjects an opportunity to engage with and shape data protection practices that govern how their personal data is protected without detriment to students' academic experience. To ensure that this data is best protected, HEIs have

to follow data protection regulations such as the GDPR [73], and have DPOs, fair use policies, and ethical codes of conduct. However, the power imbalance between students as data subjects and their institutions could weaken the data protection options available to them, particularly where not agreeing to the use of certain technologies can lead to being locked out of academic opportunities. The commons can provide data subjects with the option to understand who, what, and how their personal data is being used and ways in which they can opt-out, responding to the "Scoping and information gathering for developing the commons" section as part of our scaffolding. This includes helping them recognise their data protection rights and support them in exercising those rights without negatively affecting their ability to participate in online learning, rebalancing power between the identified stakeholder relationships. This can be done by providing templates for exercising their rights as addressed to their DPO. To facilitate opportunities to collaborate on data protection solutions within the commons, data subjects may also have the option to communicate with each other anonymously and share their insights and experiences with regards to tutorial recordings and online learning. For example, if a student is unsure how to blur their video background, they can ask for direct support from other students if instructions have not been provided in the commons. Additionally, data subjects could also be directed to the HEI's Information Technology (IT) support if they have any concerns related to recordings or online learning tools. The commons as applied to online learning can help data subjects feel more at ease about the use of online learning platforms and tools by supporting *ex ante* and *ex post* data protection solutions, providing them with an interactive, common resource to support their data protection preferences.

The online learning use case is useful for understanding how it can be applied to a data protection-focused data commons given the multitude of challenges related to the deployment of technologies and increased data gathering of students as outlined from Chapter 2.4.3. While tools such as tutorial recordings and examination monitoring can help make education more accessible and equitable, they may also hamper students' learning experiences, particularly where they are not able to opt-out of such practices. Further, there are questions as to whether these technologies are effective in their aims and objectives. By adopting online learning technologies, more student personal data is being collected, stored, analysed, and shared. As a result, students should not only be aware of the data protections and rights that they have as data subjects, but also be able to make their own decisions with regards to what their personal data protection

preferences should be. An online learning data protection-focused commons can provide data subjects with the resources to identify how their institution manages their data and what data protection rights they have, the ability to have conversations with other students or experts about any questions or concerns, and to limit the chilling effects of online learning monitoring through data protection.

5.5 Summary

In this chapter, we have introduced the data protection-focused data commons, informed by existing commons principles and theories on applying the commons to data protection. In defining a data protection-focused data commons and supporting its implementation using a data protection IAD framework and policy scaffolding, we have identified how socio-technical considerations beyond data protection compliance alone can be considered for co-creating data protection solutions. We note the following:

- By mapping commons principles, our interview findings, and the adapted IAD framework to a data protection-focused data commons framework, we have shown that a collaborative and co-created commons can support data subjects in protecting their personal data.
- We created a data protection-focused data commons policy scaffolding to identify how a commons could benefit data subjects and ensure their participation in the commons development process when applied to specific use cases.
- We have explored how a data protection-focused data commons constitutes an improvement on the state of the art in protecting personal data and can support the co-creation and collaboration of data protection solutions by data subjects through applying the commons to use cases.

Having demonstrated how a data protection-focused data commons can be theoretically applied to use cases as a socio-technical framework for data stewardship, in the next chapter, we consider how a data protection-focused can be deployed to improve the data protection process and outcome for data subjects.

CHAPTER SIX

DEPLOYING A DATA PROTECTION-FOCUSED DATA COMMONS

In Chapters 4 and 5, we established that a data protection-focused data commons can act as a socio-technical framework for data stewardship. In this chapter, we respond to our second research question posed at the beginning of this thesis on whether a data protection-focused data commons can support the practical co-creation of data protection solutions. We demonstrate how a data protection-focused data commons can be applied to the online learning use case as illustrated in Chapter 5.4.2. To do so, we conduct a user study to deploy a commons using the data protection IAD framework and investigate the usefulness of the commons for raising awareness of data protection issues, co-creating solutions from the perspective of students as data subjects.

We make the following contributions:

- Identify and apply the data protection-focused data commons to a use case.
- Build and test a data protection-focused data commons for online learning.
- Assess the extent to which a commons can support the co-creation of data protection solutions according to data subject preferences.

6.1 Method

In this chapter, we build upon the data protection-focused data commons policy scaffolding in Chapter 5.3 by translating the theoretical framework into practice through the technological deployment and assessment of a commons. This involves engaging with the policy scaffolding to address the responsibilsation, data stewardship, and data-related rights challenges illustrated in Chapter 3.2.1. We apply the scaffolding to create a data protection-focused data commons for online learning. By demonstrating the process of creating and assessing an online learning commons to support the operationalsation of a data protection-focused data commons as a pro-social socio-technical system, we address the concerns outlined in Chapter 3.2 regarding the lack of a tested commons for data protection. Further, by deploying a data protection-focused data commons, we are able test the commons' ability to act as a consensus conference to encourage dialogue among data subjects and other stakeholders in generating new knowledge together for the common good, as introduced in Chapter 3.4.

To test whether the commons can support collaborative data protection solutions, we elaborate upon our work from our case study in Chapter 2.4.3 and online learning use case application in Chapter 5.4.2. We establish three questions to examine whether an online learning data protection-focused data commons can help students regain their agency over their personal data. These are:

1. Does the ability to interact with commons resources help inform students about the purposes of online learning and tutorial recordings?

2. How effective is the commons model for supporting user preferences for protecting their personal data?

3. Does the commons model encourage more transparency around data protection between data subjects, data controllers, and other involved stakeholders?

To answer these questions, we create a commons tool, an interactive resource hub that represents a commons by applying the commons principles, that can be used by students to support them in choosing their own online learning data protection preferences. The use of the commons by students aims to help them understand the reasons behind tutorial recordings and make more informed decisions about whether they choose to consent to being recorded. The tool also attempts to

provide more agency not only on how their personal data is used by the university but also their ability to freely participate in classes.

In developing the commons, we applied Ostrom's design principles (Chapter 2) and the requirements illustrated by the IAD commons framework (Chapter 4.3) and policy scaffolding (Chapter 5.3). We also incorporated Prinsloo and Slade's learner agency framework [193]. An adapted IAD framework as applied to online learning is detailed in Appendix B. Based on the experts' perspective on how to create trust and community within a commons from Chapter 4.2, in creating the application, interactive forums and means of communicating both within the commons community as well as with external experts have been included. From the interviews, action and collective responsibility were identified as a core part of the commons. This is represented through encouraging those in the commons to share their data protection experiences with each other through the tool and including their visions of what they hope the application can help them with in the long term. Specific elements of the commons tested include building opt-in mechanisms within the platform to test whether these tools encourage data subjects to protect their personal data by making data protection-friendly choices and assessing whether having access to other data protection materials, sources, and information within a commons helps data subjects understand their data protection options. The commons also tests whether prompting data subjects to exercise their data protection rights may encourage them to learn about how their personal data is being used by data controllers. It is hoped that participation in the data protection-focused data commons encourages the redistribution of power between students as data subjects, universities as data controllers, online learning platforms, and staff.

To adapt the commons tool to online learning, we developed the application for Microsoft Teams, the software used by our university for conducting online learning. While, as noted in Chapter 2.4.3, software such as Microsoft Teams and Zoom were not developed for education, as alternatives for more open-source or privacy-friendly alternatives for online learning are to be created, developing a data protection-focused data commons on Microsoft Teams ensures greater accessibility and use by students by minimising friction between engaging with the commons and attending an online tutorial. Creating the commons on Microsoft Teams also enables HEIs to adopt the commons without needing to integrate entirely novel tools into their digital infrastructures. In our study, the commons

tool we created was then uploaded as a custom application to Microsoft Teams and each tutorial Team had a working copy of the application (Figure 6.1).

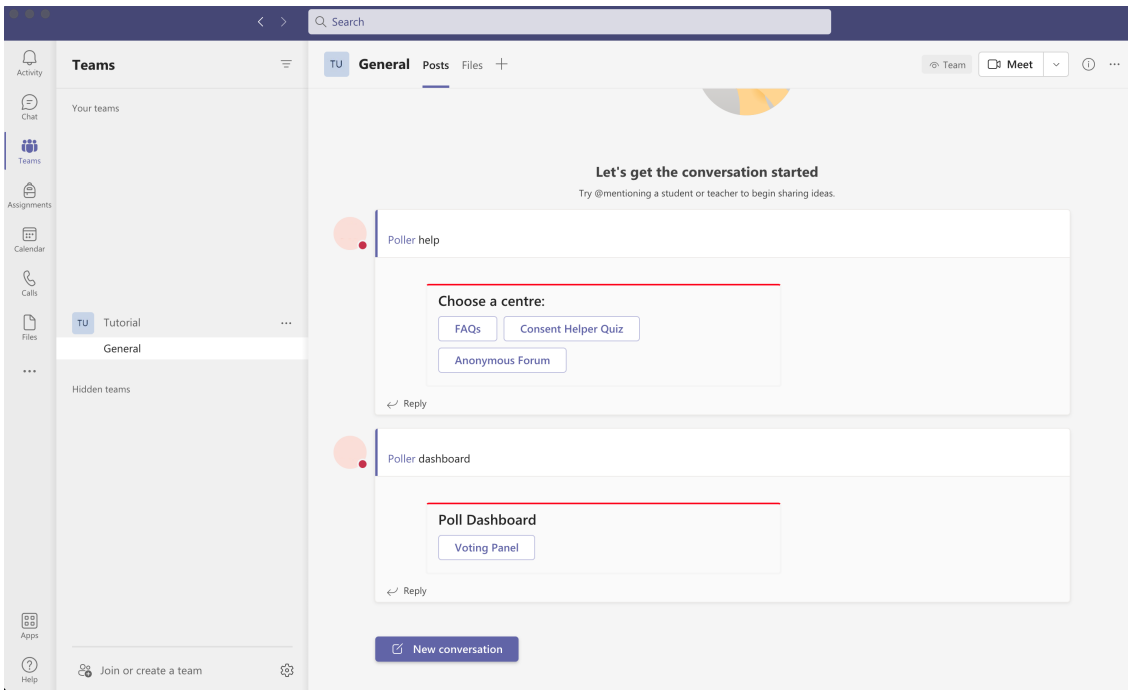


Figure 6.1: The commons tool, showing the help center and the consent voting panel, as it appears on Microsoft Teams.

The commons tool was separated into two main parts. The first part, the commons help hub (Figure 6.2), has three sections:

1. Frequently Asked Questions (FAQs) (Figure 6.3) provides answers to questions about polling, rights, policies, and contacts, mapping to the commons CPR principle for increased transparency and accountability as well as recognising the different levels of online learning governance (polycentricity).

2. Consent Helper Quiz is a short quiz to help participants figure out whether the session should be recorded, mapping to the commons CPR principle for effective management.

3. Anonymous Forum is an area for participants to share their thoughts or concerns anonymously, mapping to the commons CPR principle for citizen participation and supporting each student's equal interest.

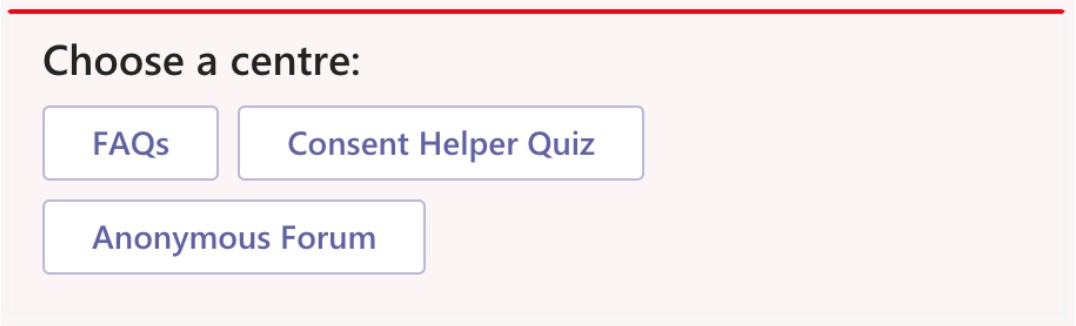


Figure 6.2: The commons tool help centre has three sections to help the student develop a more comprehensive understanding of the policies, laws, and guidance that governs tutorial recordings and supports them in making a decision as to whether or not they should consent to tutorial recording.

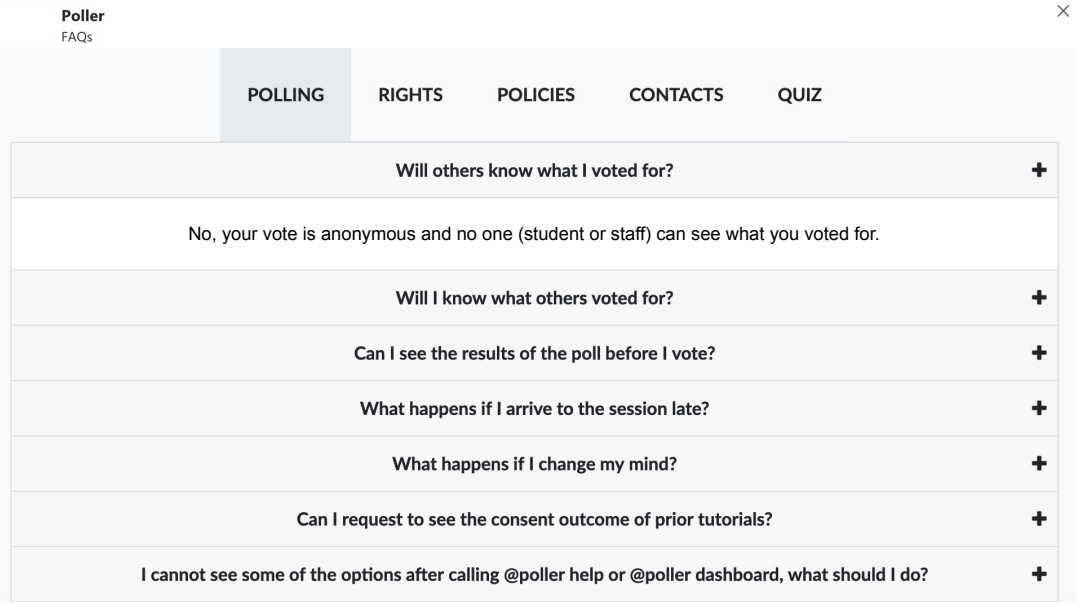


Figure 6.3: The FAQs contains text-based resources such as information about why tutorial recordings are happening, university data policies, external tutorial recording policies, and information on how to exercise data protection rights.

Within the FAQs, participants can find information about online learning, university policies, and data protection as listed below:

- Information about the tutorial recording consent Voting Panel.
- Data protection and information regulations e.g. the GDPR.
- Data protection rights centre.

  - Information about what rights data subjects (students) have.
  - Ability for students to request an anonymous record of consent poll results.
  - E-mail templates for exercising data subject rights.
  - How to contact a data protection expert and the DPO.
- Information about the data collected from the Consent Helper Quiz.

The Consent Help Quiz aims to help participants decide whether they should or should not consent to recording tutorials based on their personal preferences. All questions for the quiz have "yes" or "no" answers. Depending on the participant's answers, at the end of the quiz, the final result will display "You may not need to opt-out", "You may want to consider opting out", or "You may want to strongly consider opting out". Questions on the quiz include:

- Are you potentially revealing any sensitive personal information (racial or ethnic origin, political opinions, religious belief, genetic data, and biometric data etc.) during the session?
- Will you avoid discussing certain topics if the session is recorded?
- Will you avoid asking questions or points of clarification if the session is recorded?
- Will the session being recorded affect your likelihood of participating?
- Do you think recording the session will improve your academic study?
- Are you planning to re-watch the tutorial once it is done?
- Do you trust that the university will keep the recording safe?
- Do you trust that the platform which the session recording is taking place on will keep the recording safe?

The final part of the commons help hub is the Anonymous Forum (Figure 6.4), which allows students to share information, questions, or concerns they have about tutorial recordings.

The second part of the commons tool is the Voting Panel which conducts the consent poll (Figure 6.5).

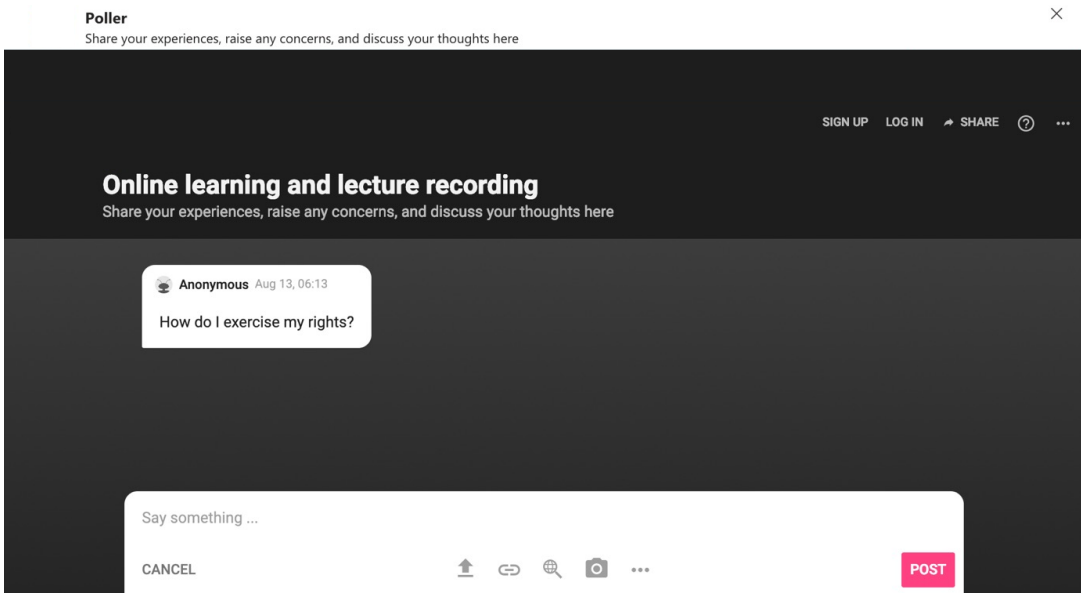


Figure 6.4: The Anonymous Forum is a space where students can participate anonymously in an open dialogue with other students in their tutorial about any questions they have about tutorial recordings.

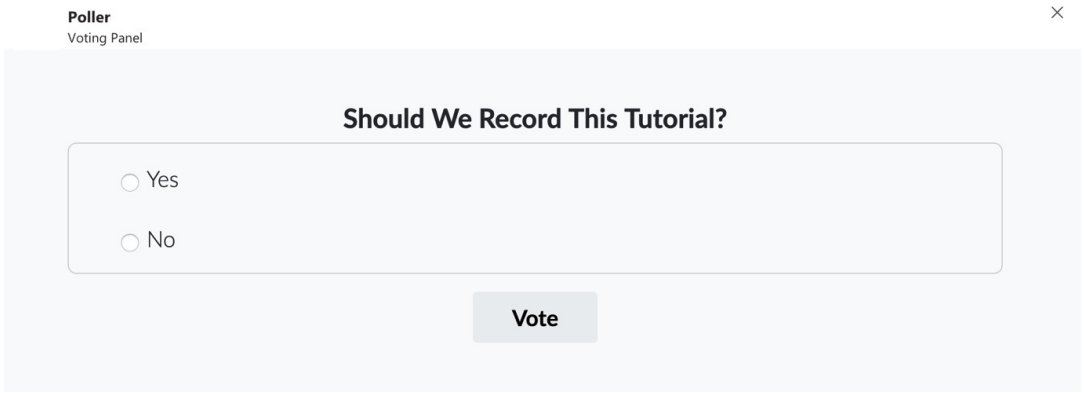


Figure 6.5: The Voting Panel is the consent poll where students can consent to or not consent to tutorial recording based on their own personal preferences. The poll is anonymous and the full results of the vote from the class will be displayed after voting. If everyone consents, the tutorial recorded.

6.1.1 Testing the application

To test the commons, we split the study into three parts: an entry questionnaire, an interactive task to test the commons or control application, and an exit questionnaire. Figure 6.6 illustrates the different stages of the study.

The entire study was conducted online between April and October 2021. All

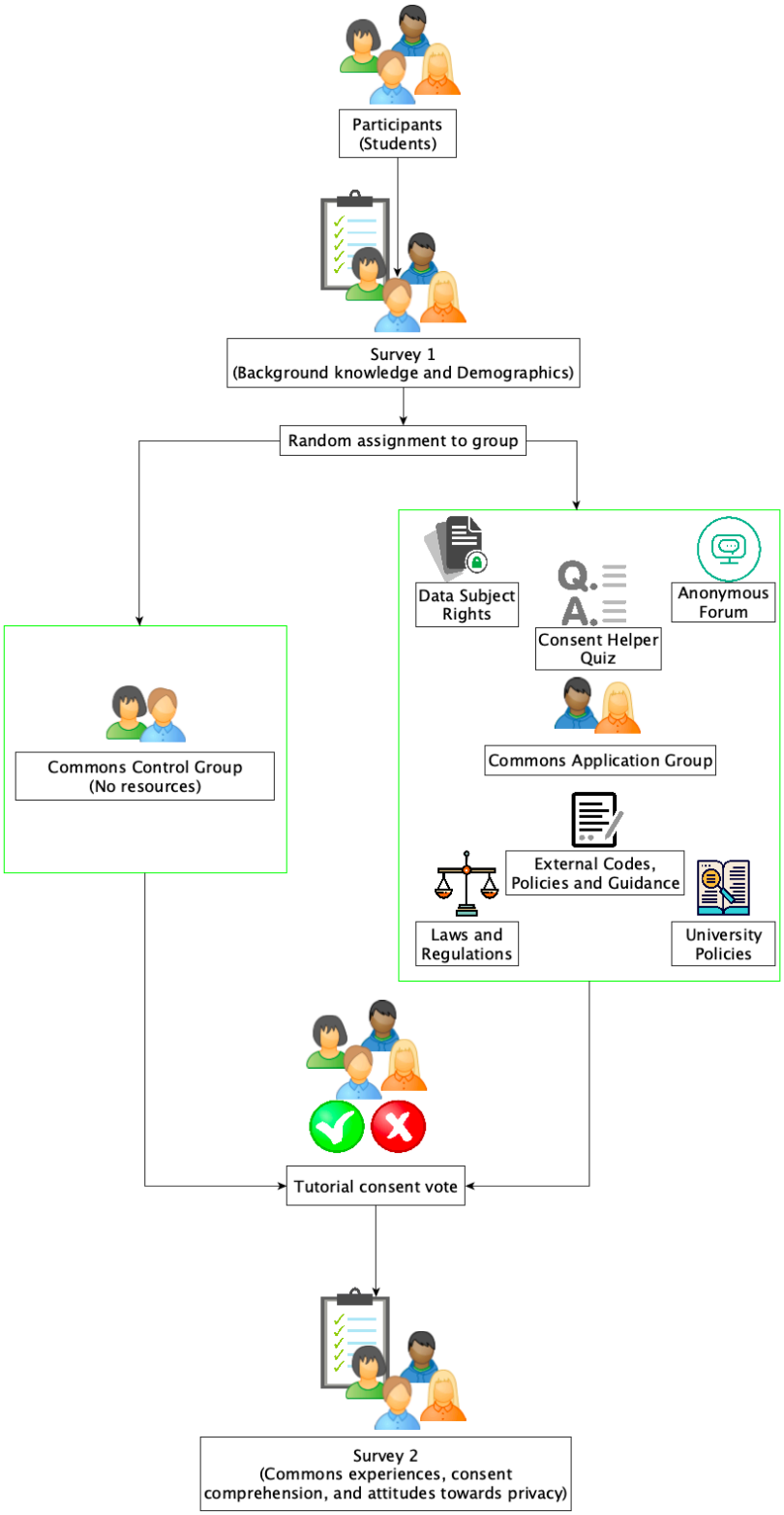


Figure 6.6: Study walk-through summary.

participants were undergraduate and postgraduate taught students studying at UK-based universities over 18 years of age.

6.1.2 Initial survey

For the first part of the study, an initial survey was completed by potential participants to gather some participant information and determine their eligibility. This assessed the level of users' knowledge of tutorial recordings, data protection, online learning, and university policies. The questionnaire also identified how participants felt about users' ability to exercise their agency with regards to tutorial recordings and online learning.

6.1.3 Testing the application

After the first survey was complete, we e-mailed potential participants to schedule a time for the rest of the study and include a separate document with the mock-tutorial information (Appendix C). The topic of the tutorial, conducting research on social media, was decided as content of the tutorial would require students to reveal personal data about themselves in addition to the recognised risks related to sharing social media data, identified in Chapters 2.4.1 and 5.4.1. Participants were then randomly assigned to be in the control testing group or the commons application testing group on Microsoft Teams. Those in the control group were given two minutes to consent or not consent to tutorial recording. Those in the commons group were given 10 minutes to explore the resources in the application and vote. The control group only had access to the voting panel and the commons application group had access to all the resources outlined in the previous section.

6.1.4 Final survey

The final part of the study, the exit survey, allowed participants to reflect on their experience of the commons, identify what resources they used if they were part of the commons application testing group, attitudes towards privacy, data protection and online learning, and examine to what extent they now know about their consent and data protection options for online learning. The survey included Internet Users' Internet Privacy Concerns (IUIPC) [147] questions adapted for online learning to benchmark their privacy concern levels that relate to privacy and data awareness, control, and collection (Appendix D).

6.2 Analysis

We recruited 34 students to participate in our study. The participants studied Computer Science (6), Management (3), Finance (2), Philosophy (2), Psychology (2), and 19 other subjects were only studied by one participant. 23 participants were undergraduates and 11 were postgraduates. Our participants predominantly identified as female (26) with seven males and one not disclosed. From our results, we did not find any correlation between the discipline of study, level or year of study, or gender.

6.2.1 Participant demographics and privacy awareness

Regarding tutorial recordings, 19 participants thought that they had control over whether a tutorial was recorded, with 10 disagreeing and five were uncertain. When asked about the university's tutorial recording policy, 17 were aware that there was one, 12 were not aware, and five were unsure. Only eight had read the policy. More broadly, most students (14) were not aware of how the university processes their personal data, 10 were unsure, and 10 were aware. Most students (22) were not aware of how Microsoft Teams processed their data.

When asked about their online learning and tutorial recording experiences, most students (20) said that some of their tutorials were recorded. 18 students said that they were asked to consent to recordings for all of their online tutorials, five said only some asked for consent, seven were not asked, and four were not sure. In considering personal experiences of online learning, 11 said that online learning made a positive impact on their educational experience, two had no impact, 12 were impacted negatively, and nine were unsure. Focusing on tutorial recordings, 13 felt that tutorial recordings were a net positive, 17 did not feel that it impacted their educational experience, two were negatively impacted, and two were unsure.

Figure 6.7 shows the overall level of privacy concern of our participants, based on their responses to online learning IUIPC questions (Appendix D). The higher the score, the more privacy-concerned a participant is, where 55 is the maximum score and 11 is the minimum. Existing work shows that internet use reduces IUIPC [20]. A positive relationship was found between privacy concerns and government involvement in privacy regulation [154], suggesting higher IUIPC scores for participants governed by the GDPR. The median score for our participants is 46. While our participants were based in the UK which falls under the GDPR's

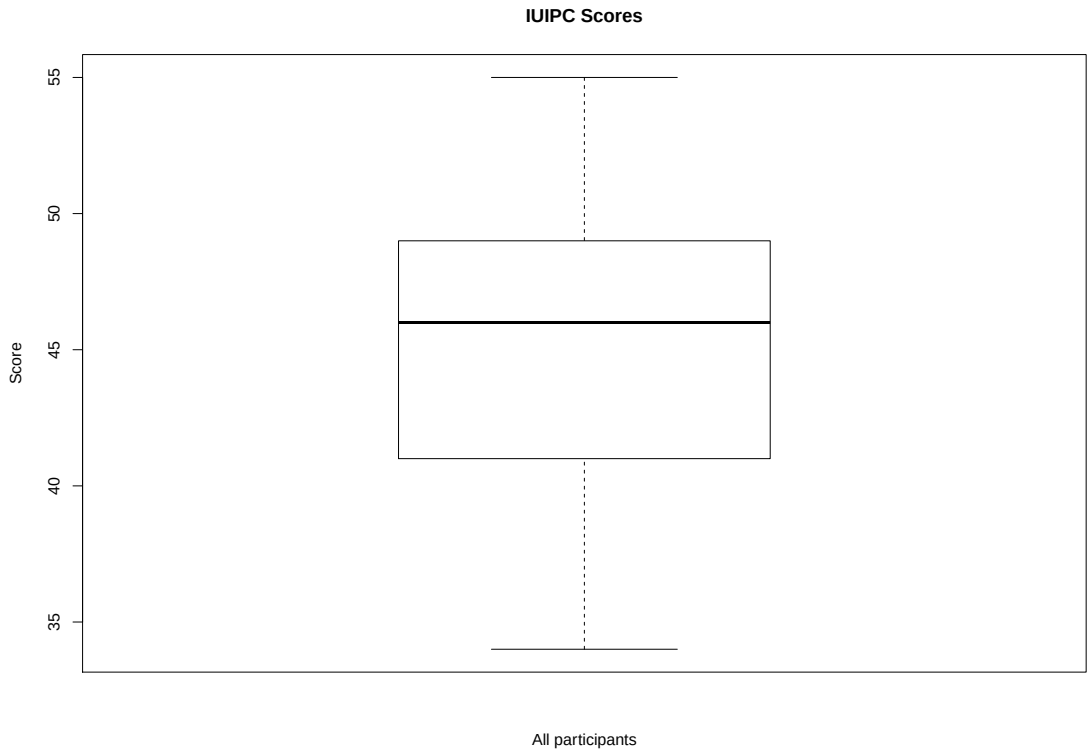


Figure 6.7: The IUIPC scores of study participants. The median score for our participants is 46, demonstrating a moderately high level of privacy concern, where 55 is the maximum score and 11 is the minimum.

remit, given that students as young people are considered to have high levels of internet use, our results suggest a relatively high level of privacy concern for their demographic. In assessing the significance of specific IUIPC questions for influencing a participant's privacy concerns, from our exploratory factor analysis (TLI of factoring reliability = 1, RMSEA index = 0, and a confidence level of 95%), we found that for data collection, participants who thought about whether they should provide personal information to universities demonstrated higher levels of privacy concern, with a correlation of 0.8. Examining the IUIPC data awareness factor, the more important participants thought it was to be aware and knowledgeable about how their personal information will be used, the higher their IUIPC score, with a correlation of 0.9. From our analysis, consent as a form of privacy control was not significant enough to be considered as a factor for assessing the level of privacy concern.

6.2.2 Consent levels for online learning

Figure 6.8 shows that most participants consented to tutorial recording. We also asked participants to state whether they decided to change how they voted as a result of doing the exit survey. One participant from the commons group and two participants from the control group would change the way they voted based on the exit survey. All three changed from not consenting to giving consent. Most students (18) stated that they did not think twice before handing over their data to the university. This suggests that students may feel obliged to provide such data to access education and indicates a certain level of trust that students have of HEIs to use that data for academic purposes.

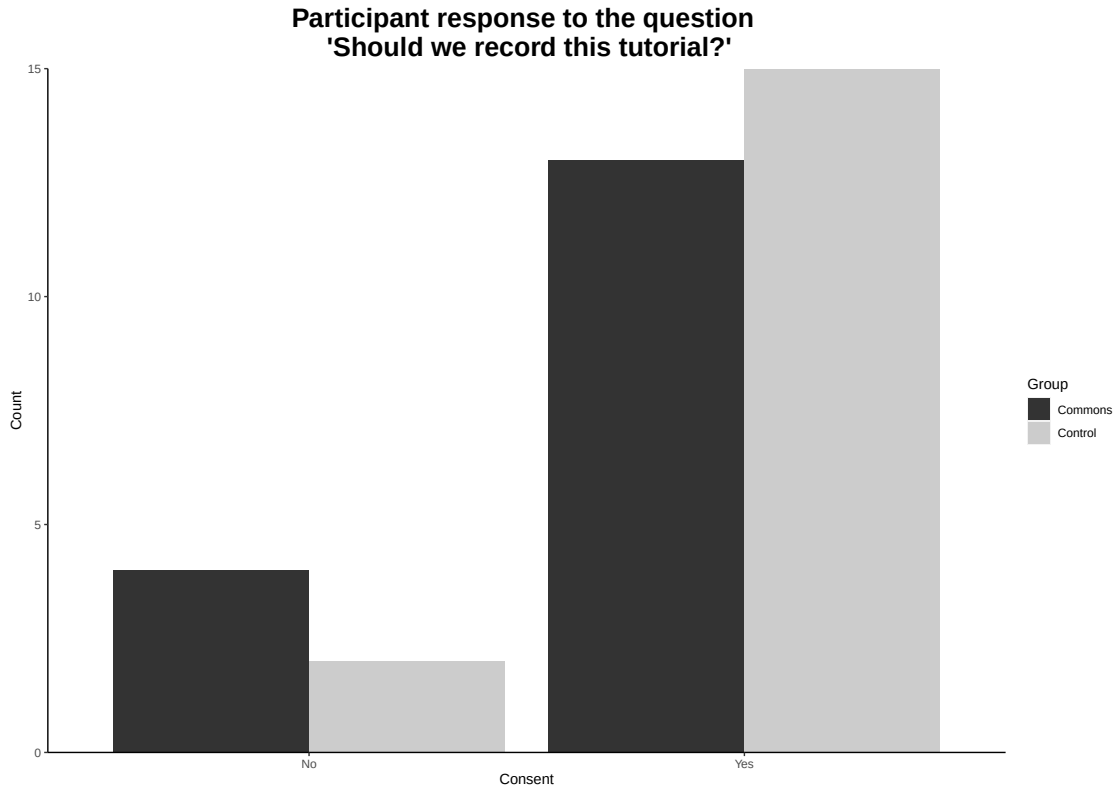


Figure 6.8: Consent preferences from participants answering the question "Should we record this tutorial?". The majority of participants in both commons and control groups consented to tutorial recording.

Several participants across both groups stated that disability and accessibility were important reasons as to why they consented to the tutorial recording. In context of the COVID-19 pandemic when the study took place, this is particularly important given the challenges students face during online learning. As a result, it is necessary to consider accessibility needs when considering whether and how tutorial recording should be conducted to support students.

6.2.3 Commons tool: information, usefulness, and agency

Table 6.1 shows that resources related to data protection rights and policies were the most useful to our study participants. This suggests that participants are keen to understand what protections are in place for their data and what recourse they have if anything goes wrong. This supports our thesis statement that the commons supports increased awareness of stakeholder tensions and develops a greater understanding of data protection options.

Commons Resource	*Sum*
Information on the University tutorial recording	13
The FAQs	13
Data protection law	12
Information on exercising your rights	12
The consent quiz	10
The consent voting poll	9
The anonymous forum	8

Table 6.1: The resources in the commons that commons group participants found useful for helping them decide whether or not they should consent to tutorial recording. All participants found at least one commons resource to be useful.

Interestingly, the actual consent voting poll where participants had to consent or not consent to the tutorial recording was the second least useful. This is consistent with existing literature on the limitations of meaningful and informed consent within [194] [158] [86] and outwith [207] [248] [30] education. This demonstrates the importance for students to feel that they are making informed choices (where the outcome is less important) in an online learning environment and may not necessarily question the university's motivations for recording tutorials.

When elaborating on why participants found certain features of the commons useful, one student said that "I hadn't really known anything about tutorial recording policy or the laws and my rights related to these recordings before so I though (*sic*) it was interesting to learn more about my rights and more about what tutorial recordings would be used for and when they should be used". Another student thought that "the information about the University policy was very valuable to make might (*sic*) decision, and having access to it easily is helpful. The FAQs was (*sic*) definitely the most helpful element, as it answered a lot of my questions simply". A student who found the forum useful explained that: "I think

I was most swayed by the anonymous student posts. Personally, I didn't want the session recorded, but I knew it would be helpful for others to review later or who had missed the session/not been mentally present due to chronic illness, etc.". Engagement with the data protection and privacy content presented in the study suggests that students do care about how their data is collected, stored, and used by the university (consistent with existing research outlined in Chapter 2), where the commons could bring this information and these concerns to light. Overall, nine participants in the commons group agreed that they would use the commons to better protect their personal data. Five somewhat agreed, two neither agreed or disagreed, and one somewhat disagreed.

6.2.3.1 Control group comparison

When the control group, where participants did not have access to the commons resources, were asked what would have be useful for them to help them decide whether or not to consent to tutorial recordings, nine participants wanted more information. These included: "Who would be able to view and access the tutorial after it had been recorded and if it would be used for anything else other than for study use for the module.", "More information on where the recording would be stored and who it would be accessible to would be helpful.", and "Whether the lecturer could see individual responses: this would influence whether I answer yes or no as I don't want to come across as a spanner in the works". From those responses, the additional information participants would have liked fell into two categories: information about the consent voting tool and information about the tutorial recording itself. Both of these are covered under the Information on the University tutorial recording section and the FAQs section of the commons. Two participants wanted to know if turning on their webcam was required as they would not consent if it was. Eight participants did not feel that they needed more information to consent either because they did not care about being recorded, would have agreed to being recorded if they knew someone in class would need the recording, or felt that they were fully aware of the tutorial recording process.

6.2.4 Topic, content, and attitudes towards tutorial recordings

When conducting the study, we asked participants to imagine that they were taking part in a mock-tutorial on conducting research on social media and provided them with the lesson plan. During their post-study survey, we asked participants

whether the topic of the mock-tutorial impacted their consent levels to tutorial recording (Figure 6.9).

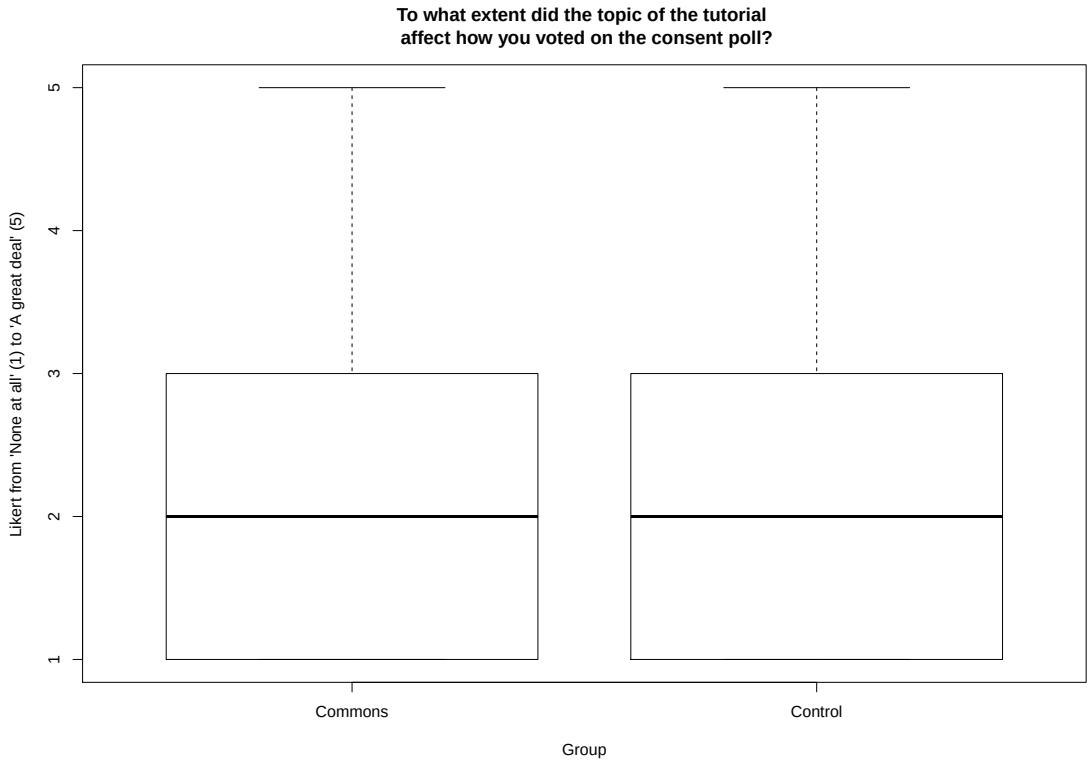


Figure 6.9: Impact of the tutorial topic on consenting to tutorial recording. The topic of the tutorial is not a strong factor for informing whether a student decides to consent or not consent to the tutorial recording in both commons and control groups. The median for both the commons and control group is two.

From the survey responses, participants suggest that they would not refuse consent based on the tutorial topic alone as it would depend on other factors such as if they felt they needed to re-watch a tutorial recording and whether the topic involves providing personal information that the participants themselves did not want to share.

We wanted to understand how students participate in recorded digital classrooms and how that may impact personal information sharing. To examine this, we asked participants whether they would avoid sharing information on any topics during online learning, specifically those listed as special category personal data under GDPR Article 9. From Figures 6.10 and 6.11, the high number of avoided topics during tutorial recordings suggest that even if participants consented to tutorial recordings, teaching subjects that result in the discussion of these sensitive personal data may limit student participation in online learning. Two of the

highest ranked topics "data concerning a person" (22) and "political opinions" (20) represent a broad range of information often shared in discussions. Six commons and three control group participants did not avoid any topic. Importantly, commons participants avoided fewer topics across all categories. This suggests that because commons participants have developed a greater understanding of how their data is stored, they are more comfortable sharing their personal data. More generally, in examining the impact of the tutorial topic and the content participants are willing to share, they explained that they would rather not participate than not consent to the tutorial recording because they had control over what they said. As a result, it is important for staff to consider how to engage with teaching sensitive topics online to maximise participation and generate the most value from online learning.

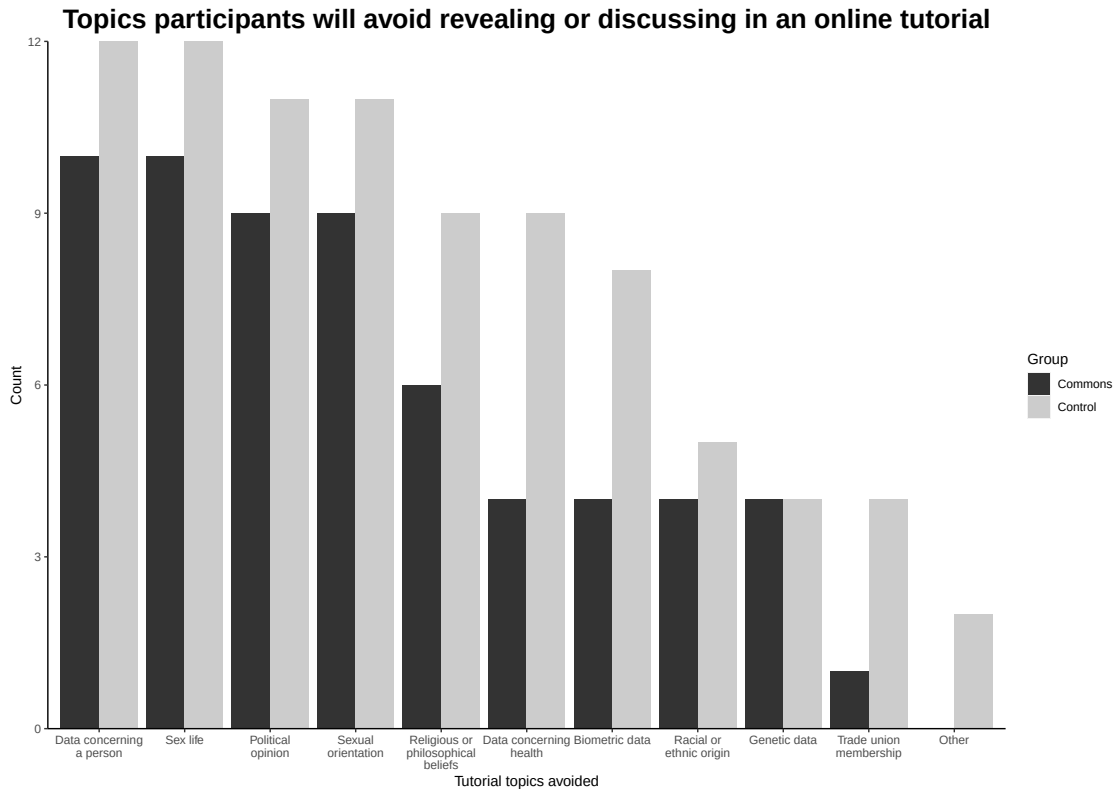


Figure 6.10: Topics participants avoided in a recorded online learning environment from participants answering the question "Are there topics you will avoid discussing or revealing about yourself if the tutorial is recorded compared to physical classes?". The two "other" responses include information that one participant considered to be "triggering" such as "mental health, other personal information, and financial information" as well as what another participant considered "anything that could be misconstrued or used out of context if the recording was inadvertently (or deliberately) released". Overall, the commons participants are less likely to avoid discussing certain topics.

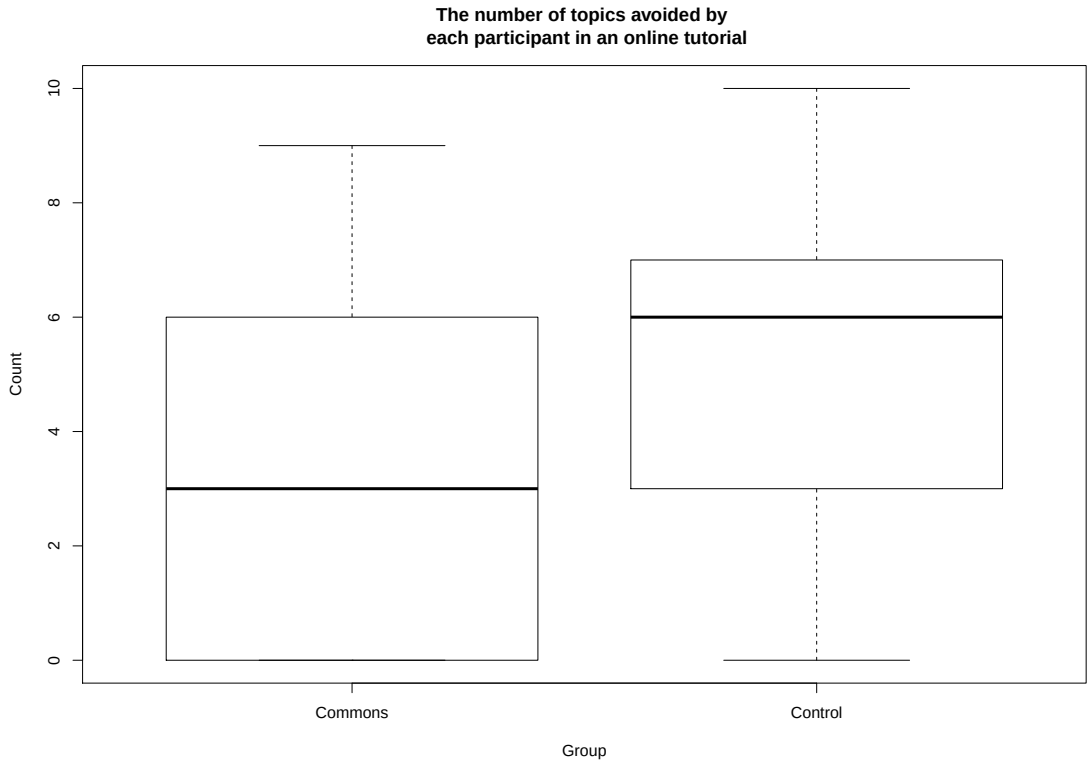


Figure 6.11: The number of topics avoided by each participant. Commons participants are less likely to avoid discussing certain topics during online learning, where the median for topics avoided is three compared to six from the control group.

6.2.5 Summary

In response to our study questions from the beginning of this chapter, we find that interacting with commons resources helps inform students about the purpose of online learning and tutorial recordings. From our findings across both groups, students found the commons useful in supporting their data protection preferences. Some students in the control group would also have liked more information about how their data was being collected and used when being recorded.

Responding to our second thesis research question on whether a data protection-focused data commons can support the co-creation of data protection solutions for the benefit of data subjects, we found that the commons model is useful for supporting user preferences for protecting their personal data because it helps students learn more about how their data is collected, used, and stored. Almost all participants consented to tutorial recording, indicating that students find value, both for themselves and the class overall, when it comes to being able to access a recording. Most students also indicated factors such as accessibility and helping

other students as reasons for consenting to recordings. While recording the level of consent can be useful in understanding whether participants feel comfortable with the collection of data for tutorial recordings, it only provides a limited picture as to the extent to which the ability to interact with the commons helps inform them about the purposes of online learning and tutorial recording. Additionally, with the commons prompting students to consider their consent preferences as well as present resources related to tutorial recording and data protection, the voting panel as part of the commons itself can be presented as a boundary object [163] to encourage collaboration and engagement with the commons. Therefore, it is important to consider other factors and means for understanding student agency and collaborative solution-building for supporting their data protection preferences.

More transparency around data protection between students, staff, university management, and other involved stakeholders is encouraged through a commons as the model supports the identification of stakeholder tensions and breaks them down through identifying a common aim — accessing a dynamic, participatory learning environment. The high levels of consent to tutorial recording could imply that students, to some extent, trust universities with their online learning data. However, given that more information on data protection was found to be preferable, students may want more details on how and what data is collected and used. This is particularly true in preventing potential harm should there be any data breaches, given the higher preference for understanding data protection and data subject rights. The commons encourages more transparency around data protection between students and other stakeholders, increasing awareness of different stakeholder interest through its polycentric governance. This is because the commons not only informs students of the data governance and risk management polices related to online learning data, but also supports recourse through data protection rights if any harms are realised.

More broadly, the commons can be useful for reconsidering online learning socio-technical pedagogies in context of data protection to support more inclusive and safe digital classrooms. Our results indicate that when asked about students' participation in recorded online tutorials compared to in-person sessions, most participants indicated that there would be topics that they would avoid discussing. This suggests that while students are happy to consent to tutorial recordings, they may decrease their level of participation in online classes. This could impact the

quality of tutorial participation in online teaching. As a result, staff could be mindful of asking students questions related to their personal experience that may reveal these forms of data. Commons participants are also more willing to reveal their personal data, suggesting that developing an understanding of what and how data is collected and processed can encourage participation. Overall, staff and academic institutions may want to consider how the online learning environment could be fostered to maintain the privacy and security offered by the physical classroom.

6.3 Summary

In this chapter, we set out to build and deploy a data protection-focused data commons for online learning to test whether a data protection-focused data commons can support the co-creation of data protection solutions for the benefit of the data subject. To test this, we created a commons by applying the data protection IAD framework to online learning and establishing other social-technical requirements for the commons development. We note the following:

- In a user study of 34 student participants, we determined that the commons can support student preferences for protecting their personal data both *ex ante* and *ex post*. Greater transparency between students, university management, and use of data by online learning platforms can help students feel more assured about how their data is used and what recourse may be available if they have any data protection concerns.
- Consent as a means for informing students about tutorial recording is insufficient, where more clarity on university data protection policies and data subject rights can help support student agency with regards to their personal data.
- The sharing of information and knowledge through the interactive components of the online learning commons allows data subjects to collaborate on and discuss data protection solutions without detriment to their academic progress.
- The governance of online learning data through a commons supports protection beyond data protection law, as there are wider socio-technical, ethical, and wellbeing considerations on how EdTech should be deployed.

Having established the data protection-focused data commons as an appropriate socio-technical framework for protecting personal data and supporting the co-creation of data protection solutions, in the next chapter, we consolidate and discuss the implications of the findings we have made in this thesis, and outline directions for further work.

CHAPTER SEVEN

CONCLUSION

Our data-driven society has resulted in the increased collection, management, and sharing of data subjects' personal data by large, international data controllers. The power imbalance between them has led to individuals not being able to understand, control, mitigate, or seek recourse for their data rights and any potential harms when it comes to their personal data. While there are legal and technological solutions that aim to address this, the data protection process has become responsibilised, where data subjects have to find data protection solutions themselves. To tackle the responsibilisation of the personal data process and rebalance power between data subjects and data controllers, we assessed whether commons theories and principles can be applied to data protection to support the co-creation of data protection solutions for the benefit of the data subject as a socio-technical framework for data stewardship. First, we interviewed commons experts to understand the practical considerations of creating a commons and how data protection had been previously applied. Next, we define the data protection-focused data commons and applied an adapted data protection IAD framework and policy scaffolding to potential use cases. Finally, we applied the online learning use case to the deployment of a data protection-focused data commons to test whether the commons can support the co-creation of data protection solutions from the perspective of students as data subjects. In summary, we have addressed the following thesis:

> A data commons created with the aim of protecting personal data can encourage data subjects to co-create and collaborate on data protection solutions, increasing awareness of different stakeholder interests as enabled by data protection law.

7.1 Contributions

To test our thesis statement, we considered the following questions:

RQ1: Is a data protection-focused data commons appropriate as a socio-technical framework for data stewardship?

RQ2: Can a data protection-focused data commons support the co-creation of data protection solutions for the benefit of data subjects?

To address the first question, in Chapter 4 we conducted interviews with commons experts to determine how the commons could be improved and adapted for data protection. In Chapter 5, we applied data protection principles to the commons framework to create a data protection-focused data commons, and illustrated potential use cases.

In Chapter 4, we identified the benefits and challenges to creating a commons through interviewing experts. The findings from our interviews informed how a commons could be adapted to create a data protection-focused data commons focused on increasing data subject engagement, participation, and co-creation, which in Chapter 3 we found was not always considered in the state of the art. We demonstrated how the IAD framework could be adapted to support the practical implementation of a data protection-focused data commons to support data subjects in expressing their data protection preferences.

In Chapter 5, we introduced the data protection-focused data commons in more detail. Building upon existing theories, methodologies, and practical applications of the commons, we extended the commons beyond the dissemination of data to create a data protection-focused data commons for data subjects to further their ability to protect the processing of their personal data. We adapt a data protection policy scaffolding to illustrate how this can be achieved and what it could look like in practice by applying the commons to potential use cases.

To address the second question, in Chapter 6, we built and deployed a data protection-focused data commons for online learning to examine the usefulness of the commons from the perspective of data subjects. In a user study that examined if and how a commons could support data subjects' data protection preferences in a collaborative digital environment, we found that the commons can support student preferences for protecting their personal data both *ex ante* and *ex post*, where greater transparency between students, university management, and use

of data by online learning platforms can help students feel more assured about how their data is used and what recourse may be available if they have any data protection concerns. Consent as a means for informing students about tutorial recording is insufficient, and the protection and governance of online learning data should go beyond data protection, as there are wider ethical and well-being considerations on how education technologies should be deployed.

7.2 Discussion and future work

In this thesis, we have demonstrated how the adaptation of the commons, IAD framework, and policy scaffolding can support data protection practices, where the co-creation, collaboration, and engagement from data subjects can help them express their personal data preferences. As we have argued, this is an important advance on the state of the art, addressing the power imbalance between data subjects and data controllers as well as the responsibilisation of the data protection process. We do acknowledge, however, that the data protection-focused data commons is not a one-size-fits-all solution to data protection and data governance challenges. The data protection-focused data commons can aid data subjects to collectively understand the data protection landscape and express their individual and group personal data protection preferences. However, the data protection-focused data commons is not able to challenge the wider socio-technical and economic infrastructures and systems that enable the vast collection, management, and sharing of data. This is an acknowledged problem within the research area for data governance that is outwith the scope of this work. In Chapter 3.2, we discussed recent and ongoing multidisciplinary work to address the challenges posed by the responsibilisation of the data protection process that could address this issue. For example, the Data Trust Initiative is establishing pilots to test the real-world application of data trusts as a legal mechanism.

7.2.1 Empirical research

In Chapter 6, we deployed a data protection-focused data commons to assess the usefulness of the commons for co-creation and collaboration. Our study demonstrated that a commons for online learning can support the protection of personal data and encourage the co-creation of data protection solutions through increased awareness of different stakeholder tensions, providing greater transparency on data regulations and the means to exercise their data protection

rights, but there are limitations to our empirical work. Regarding our study, those who opted to participate are likely to be more privacy aware. Several participants mentioned that because data protection and tutorial recordings were mentioned in the study description, the thought was already on their minds beforehand. For unknown reasons, more students who identified as female participated in our study. Although we did not find any patterns or correlation to gender in response to our surveys, greater gender balance may be preferred to mitigate any potential biases. Additionally, it was initially hoped that the study could have been done in groups to more accurately mimic the tutorial environment. However, challenges in recruiting participants, time zone differences, and asking them to spend more time on Microsoft Teams outside of classes (as an indication, 175 participants completed the initial survey but only 34 responded to the following Microsoft Teams study) meant that it was difficult to schedule participants to the same session. As a result, 28 students participated in the study individually and three pairs participated together. There was no identifiable difference between their responses.

Given that our online learning commons was only tested on students, further research could be done with staff to examine whether the commons could be useful for protecting their agency for protecting personal data. This is particularly important due to concerns of HEIs using EdTech to monitor staff [11] and break union strikes [237], where intellectual property rights do not always belong to the individual who produced the work [255] [148]. With the rise of children's data collection in the classroom, the commons could also be tested on younger learners to examine its usefulness for students, teachers, and parents. Lastly, we acknowledge that online learning during the COVID-19 pandemic when the study took place is different to what it might have been if technologies were implemented more organically. Students, staff, and universities have had to instantly adapt to shifting the physical classroom into a digital one. As a result, future work assessing the hybrid learning education landscape post-pandemic may be beneficial for assessing how data subjects' attitudes and HEIs' policies may have changed with respect to data protection.

7.2.2 Socio-technical developments for a commons

As we acknowledged in Chapter 4, the use of the commons as part of the data protection and governance process is a choice from the perspectives of users and data subjects. As a result, we consider that the data protection-focused data

commons is a solution that can provide data subjects with another data protection choice that focuses specifically on their engagement and collaboration as part of the data protection process. Even if other data stewardship and data governance mechanisms remain available, the data protection-focused data commons is a valuable and practical means in which data protection solutions can be co-created, bridging the gap between law and policy-based developments with socio-technical applications within existing institutions.

Further, we have discussed the data protection-focused data commons' contribution to improving the state of protecting personal data through co-created and collaborative means. The design of the data protection-focused data commons has been informed by grounding it in Ostrom's work on the commons, and is shaped by the adaptations for data protection-related laws and technologies detailed in Chapter 2. This body of work has identified useful extensions to the commons, demonstrated in Chapter 3. In Chapter 5, we proposed the data protection-focused data commons to encourage more awareness on data protection and the use of data subjects' own personal data, allowing them to make choices that are more in line with their own preferences. The creation of a data protection-focused data commons could be further shaped by practical implementation of data governance mechanisms in the future. While our approach allows organisations and institutions to adopt data protection solutions through data subject engagement, further work is needed to address concerns about data protection, governance, and management more broadly. More generally, the data protection-focused data commons could be expanded to actively incorporate other stakeholders in seeking greater redress to tensions between them and data subjects.

To deploy a data commons in the long-term, considerations need to be made with regards to the platform used to host the commons and how the commons is to be sustained financially. This includes technical infrastructure and system considerations, particularly whether the commons could be created within an existing digital ecosystem or built independently. These decisions should be made in consultation with data subjects based on their accessibility, data and data protection requirements, as well as expert advice. Given the difference in stakeholder interests, how and by whom the commons is maintained can impact the trust between users as data subjects and others participating in the commons' development. As the focus is still placed on data subject preferences, this can more clearly draw out where the power imbalances lie and how those

data protection challenges can be more equitably addressed. Therefore, even if new data governance mechanisms emerge, the data protection-focused data commons can still facilitate meaningful conversations to identify what individuals and collective data subjects want when it comes to agency over their personal data.

7.2.3 Law and policy

Wider conversations between data protection stakeholders, identified in Chapter 2, could be facilitated to raise awareness on how personal data is being treated in our data-driven society within a data-protection focused data commons from legal, socio-technological, and ethical perspectives as supported by our interview findings in Chapter 4.2. This includes discussing the impact of data-related regulations and policies on data subjects. For example, within data protection regulations, access to the fundamental right to data protection through the exercise of data rights can be further strengthened [9]. Laws such as the European Data Governance Act, which is drafted and currently subject to the European Council's approval, aims to increase trust in data intermediaries and strengthen data-sharing mechanisms across the EU, could support broader data protection practices for the benefit of data subjects outside of data protection [75]. Research and guidance from organisations and advisory bodies such as the Centre for Data Ethics and Innovation in the UK can play an important role connecting different stakeholders and addressing data issues to specific domains, including the data infrastructures needed to support a commons [35]. By encouraging discussions around data governance beyond data protection, a data protection-focused data commons can not only address data subject issues but also take into account the bigger picture in relation to how personal data can be protected for the common good.

7.3 Summary

In this thesis, we have shown that a data protection-focused data commons can encourage data subjects to co-create data protection solutions through engagement and collaboration. From identifying the benefits and challenges to creating a commons, through to adapting the commons to data protection, and deploying and analysing the usefulness of the commons, a data protection-focused data commons has provided an alternative socio-technical solution that supports data subject agency as part of the data protection process.

References

[1] Aaron Ansuini (@AaronLinguini). 2021. Twitter post: HI EXCUSE ME, I just found out the the prof for this online course I'm taking *died in 2019* and he's technically still giving classes since he's *literally my prof for this course* and I'm learning from lectures recorded before his passing .....it's a great class but WHAT. https://twitter.com/AaronLinguini/status/1352009211501289472. Accessed: 2021-09-20.

[2] ACLU. 2013. You are being tracked. https://www.aclu.org/issues/privacy-technology/location-tracking/you-are-being-tracked

[3] Ada Lovelace Institute. 2021. Data Cooperatives. https://www.adalovelaceinstitute.org/feature/data-cooperatives/

[4] Ada Lovelace Institute. 2021. Data Trusts. https://www.adalovelaceinstitute.org/feature/data-trusts/

[5] Ada Lovelace Institute. 2021. Exploring legal mechanisms for data stewardship. https://www.adalovelaceinstitute.org/wp-content/uploads/2021/03/Legal-mechanisms-for-data-stewardship_report_Ada_AI-Council-2.pdf

[6] Ada Lovelace Institute. 2021. Exploring principles for data stewardship - a case study analysis. https://docs.google.com/spreadsheets/d/1hAN8xMJuxobjARAWprZjtcZgq1lwOiFT7hf2UsiRBYU/edit#gid=432908716

[7] Ada Lovelace Institute. 2021. Participatory data stewardship. https://www.adalovelaceinstitute.org/report/participatory-data-stewardship

[8] Ida-Elisabeth Andersen and Birgit Jæger. 1999. Scenario workshops and consensus conferences: Towards more democratic decision-making. *Science*

and Public Policy 26, 5 (10 1999), 331–340. https://doi.org/10.3152/147154399781782301

[9] Jef Ausloos, Réne Mahieu, and Michael Veale. 2020. Getting Data Subject Rights Right A submission to the European Data Protection Board from international data rights academics, to inform regulatory guidance. *JIPITEC* 10, 3 (2020), 283–309. https://nbn-resolving.de/urn:nbn:de:0009-29-50315

[10] Australia Research Data Commons. 2020. Australia Research Data Commons. https://ardc.edu.au

[11] Evronia Azer. 2021. Remote working has led to managers spying more on staff – here are three ways to curb it. https://theconversation.com/remote-working-has-led-to-managers-spying-more-on-staff-here-are-three-ways-to-curb-it-159604

[12] Kirstie Ball and Laureen Snider. 2013. *The surveillance-industrial complex: A political economy of surveillance*. Taylor and Francis, London, UK. https://doi.org/10.4324/9780203094426

[13] Kirstie Ball and William Webster. 2020. Big Data and surveillance: hype, commercial logics and new intimate spheres. *Big Data & Society* 7, 1 (14 May 2020), 1–5. https://doi.org/10.1177/2053951720925853

[14] Solon Barocas and Karen Levy. 2020. Privacy dependencies. *Washington Law Review* 95, 2 (2020), 555–616. https://digitalcommons.law.uw.edu/wlr/vol95/iss2/4

[15] Lindsey Barrett. 2021. Rejecting Test Surveillance in Higher Education. *Michigan State Law Review (forthcoming)* 1, 1 (2021), 83. https://doi.org/10.2139/ssrn.3871423

[16] Siân Bayne, Peter Evans, Rory Ewins, Jeremy Knox, James Lamb, Hamish Macleod, Clara O'Shea, Jen Ross, Philippa Sheail, and Christine Sinclair. 2020. *The Manifesto for Teaching Online*. MIT Press, London, UK.

[17] BBC. 2017. Chatbot offers legal help to Equifax data breach victims. https://www.bbc.co.uk/news/technology-41239513

[18] BBC. 2020. 'Zoombombing' targeted with new version of app. https://www.bbc.co.uk/news/business-52392084

[19] BBC. 2021. Hong Kong: How life has changed under China's national security laws. https://www.bbc.co.uk/news/world-asia-china-57649442

[20] Steven Bellman, Eric Johnson, Stephen Kobrin, and Gerald Lohse. 2004. International Differences in Information Privacy Concerns: A Global Survey of Consumers. *Inf. Soc.* 20 (11 2004), 313–324. https://doi.org/10.1080/01972240490507956

[21] Bennett Institute for Public Policy and the Open Data Institute. 2020. The Value of Data Summary Report 2020. https://www.bennettinstitute.cam.ac.uk/media/uploads/files/Value_of_data_summary_report_26_Feb.pdf

[22] Jarg Bergold and Stefan Thomas. 2012. Participatory Research Methods: A Methodological Approach in Motion. *Forum Qualitative Sozialforschung* 13, 1 (2012), 31. https://nbn-resolving.de/urn:nbn:de:0114-fqs1201302

[23] Bits of Freedom. 2016. My Data Done Right. https://mydatadoneright.eu/

[24] Irina Bolychevsky. 2018. How solid is Tim's plan to redecentralize the web? https://medium.com/zero-equals-false/how-solid-is-tims-plan-to-redecentralize-the-web-b163ba78e835

[25] Coline Boniface, Imane Fouad, Nataliia Bielova, Cédric Lauradoux, and Cristiana Santos. 2019. Security Analysis of Subject Access Request Procedures How to authenticate data subjects safely when they request for their data. In *APF 2019 - Annual Privacy Forum*. HAL, INRIA, Rome, Italy, 1–20. https://hal.inria.fr/hal-02072302

[26] Christine L. Borgman. 2018. Open Data, Grey Data, and Stewardship: Universities at the Privacy Frontier. *Berkeley Technology Law Journal* 33, 2 (2018), 365–412. https://doi.org/10.15779/Z38B56D489

[27] Sarah Boseley. 2016. NHS to scrap single database of patients' medical details. https://www.theguardian.com/technology/2016/jul/06/nhs-to-scrap-single-database-of-patients-medical-details

[28] Francesca Bria. 2021. The EU must be bold and defend its digital sovereignty. https://www.ft.com/content/84dbe3a0-3a40-43bd-850d-ba8e3cab34cd

[29] Chris Brook. 2021. 2021 to Date Has Seen More Data Breaches Than 2020. https://digitalguardian.com/blog/2021-date-has-seen-more-data-breaches-2020.

[30] Lee A. Bygrave and Dag Wiese Schartum. 2009. Consent, Proportionality and Collective Power. In *Reinventing Data Protection?*, Serge Gutwirth, Yves Poullet, Paul De Hert, Cécile de Terwangne, and Sjaak Nouwt (Eds.). Springer Netherlands, Dordrecht, The Netherlands, 157–173.

[31] Carole Cadwalladr and Emma Graham-Harrison. 2018. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election

[32] California State Legislature. 2018. The California Consumer Privacy Act of 2018. *California Legislative Information* 375 (2018), 1–24. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180AB375

[33] Leyland Cecco. 2020. Google affiliate Sidewalk Labs abruptly abandons Toronto smart city project. https://www.theguardian.com/technology/2020/may/07/google-sidewalk-labs-toronto-smart-city-abandoned

[34] Center for Democracy & Technology. 2021. Navigating the New Normal: Ensuring Equitable and Trustworthy EdTech for the Future. https://cdt.org/wp-content/uploads/2021/11/CDT-Teacher-Parent-Student-Survey-Fall-2021-Final.pdf

[35] Centre for Data Ethics and Innovation. 2014. About Us. https://www.gov.uk/government/organisations/centre-for-data-ethics-and-innovation/about

[36] Monica Chin. 2020. An ed-tech specialist spoke out about remote testing software — and now he's being sued. https://www.theverge.com/2020/10/22/21526792/proctorio-online-test-proctoring-lawsuit-universities-students-coronavirus

[37] Shruthi Sai Chivukula, Chris Rhys Watkins, Rhea Manocha, Jingle Chen, and Colin M. Gray. 2020. Dimensions of UX Practice That Shape Ethical Awareness. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, USA, 1–13. https://doi.org/10.1145/3313831.3376459

[38] Roger Clarke. 1988. Information Technology and Dataveillance. *Commun. ACM* 31, 5 (May 1988), 498–512. https://doi.org/10.1145/42411.42413

[39] Care Quality Commission. 2016. Safe data, safe care. https://www.cqc.org.uk/publications/themed-work/safe-data-safe-care

[40] European Commission, Joint Research Centre, and Kirstie Ball. 2021. *Electronic monitoring and surveillance in the workplace : literature review and policy recommendations*. European Commission, European Union. https://doi.org/10.2760/451453

[41] Andrew Cormack. 2016. A Data Protection Framework for Learning Analytics. *Journal of Learning Analytics* 3 (2016), 91–106. https://doi.org/10.18608/jla.2016.31.6

[42] Nick Couldry and Alison B. Powell. 2014. Big Data from the bottom up. *Big Data & Society* 1, 2 (2014), 2053951714539277. https://doi.org/10.1177/2053951714539277

[43] Martin Coulter. 2021. Alphabet's Sidewalk Labs has abandoned another US smart city project after reported fights about transparency. https://www.businessinsider.com/second-sidewalk-labs-smart-city-project-shutters-portland-oregon-2021-2?r=US&IR=T

[44] Diane Coyle. 2020. Common governance of data: appropriate models for collective and individual rights. https://www.adalovelaceinstitute.org/blog/common-governance-of-data/

[45] Andy Crabtree, Tom Lodge, James Colley, Chris Greenhalgh, Richard Mortier, and Hamed Haddadi. 2016. Enabling the new economic actor: data protection, the digital economy, and the Databox. *Pers Ubiquit Comput* 20, 1 (2016), 947–957. https://doi.org/10.1007/s00779-016-0939-3

[46] Bart Custers, Alan M. Sears, Francien Dechesne, Ilina Georgieva, Tommaso Tani, and Simone van der Hof. 2019. *Conclusions*. T.M.C. Asser Press, The Hague, The Netherlands, 195–233. https://doi.org/10.1007/978-94-6265-282-8_10

[47] Data Economy Lab. 2021. Data Cooperative. https://tool.thedataeconomylab.com/data-models/1

[48] Data Economy Lab. 2021. Data Stewardship models. https://tool.thedataeconomylab.com/our-data-models

[49] Data Economy Lab. 2021. Data Trust. https://tool.thedataeconomylab.com/data-models/10

[50] Data Trusts Initative. 2021. Seeking Data Trusts Pioneers! Funding from the Data Trusts Initiative will support pilot projects to set up real-world data trusts. https://datatrusts.uk/blogs/seeking-data-trusts-pioneers-funding-from-the-data-trusts-initiative-will-support-pilot-projects-to-set-up-real-world-data-trusts

[51] Datastreams. 2017. Datastreams homepage. https://www.datastreams.io/

[52] Paul De Hert, Vagelis Papakonstantinou, Gianclaudio Malgieri, Laurent Beslay, and Ignacio Sanchez. 2018. The right to data portability in the GDPR: Towards user-centric interoperability of digital services. *Computer Law & Security Review* 34, 2 (April 2018), 193–203. https://doi.org/10.1016/j.clsr.2017.10.003

[53] Bruna De Marchi. 2003. Public participation and risk governance. *Science and Public Policy* 30, 3 (06 2003), 171–176. https://doi.org/10.3152/147154303781780434

[54] Decidim. 2021. Decidim homepage. https://decidim.org/

[55] DECODE. 2020. DECODE Application. https://github.com/DECODEproject/decode-app

[56] Sylvie Delacroix and Neil D Lawrence. 2019. Bottom-up data Trusts: disturbing the 'one size fits all' approach to data governance. *International Data Privacy Law* 9 (2019), 236–252. https://doi.org/10.1093/idpl/ipz014

[57] Mary Dellenbaugh-Losse, Nils-Eyk Zimmermann, and Nicole de Vries. 2020. *The Urban Commons Cookbook: Strategies and Insights for Creating and Maintaining Urban Commons*. IngramSpark, La Vergne, USA.

[58] Josh Denny, David Glazer, Robert L. Grossman, Benedict Paten, and Anthony Philippakis. 2017. A Data Biosphere for Biomedical Research. https://medium.com/@benedictpaten/a-data-biosphere-for-biomedical-research-d212bbfae95d

[59] Department for Digital, Culture, Media & Sport and The Rt Hon Oliver Dowden CBE MP. 2021. UK unveils post-Brexit global data plans to boost growth, increase trade and improve healthcare. https://www.gov.uk/government/news/uk-unveils-post-brexit-global-data-plans-to-boost-growth-increase-trade-and-improve-healthcare

[60] Ada Diaconescu and Jeremy Pitt. 2017. Technological Impacts in Socio-Technical Communities: Values and Pathologies. *IEEE Technology and Society Magazine* 36, 3 (2017), 63–71. https://doi.org/10.1109/MTS.2017.2728780

[61] Claudia Diaz, Omer Tene, and Seda F. Guerses. 2013. Hero or Villain: The Data Controller in Privacy Law and Technologies. *Ohio State Law Journal* 74, 6 (2013), 923–964.

[62] DoNotPay. 2019. DoNotPay homepage. https://https://donotpay.com/

[63] Paul Dourish and Ken Anderson. 2006. Collective Information Practice: Emploring Privacy and Security as Social and Cultural Phenomena. *Hum.-Comput. Interact.* 21, 3 (9 2006), 319–342. https://doi.org/10.1207/s15327051hci2103_2

[64] Joseph Duball. 2020. Shift to online learning ignites student privacy concerns. https://iapp.org/news/a/shift-to-online-learning-ignites-student-privacy-concerns/

[65] Lilian Edwards. 2018. Data Protection: Enter the General Data Protection Regulation. In *Law, Policy and the Internet*, Lilian Edwards (Ed.). Hart Publishing, London, UK, Chapter 4, 77–117.

[66] Erin Egan and Ashlie Beringer. 2018. It's Time to Make Our Privacy Tools Easier to Find. https://newsroom.fb.com/news/2018/03/privacy-shortcuts/

[67] NHS England. 2013. NHS England sets out the next steps of public awareness about care.data. https://www.england.nhs.uk/2013/10/care-data/

[68] European Commission. 2018. Reclaiming the Smart City: Personal data, trust and the new commons. *Decode* 1, 1 (2018), 77. https://media.nesta.org.uk/documents/DECODE-2018_report-smart-cities.pdf

[69] European Commission. 2019. European Open Science Cloud (EOSC) strategic implementation plan. *European Commission* 01 (07 2019), 95. https://op.europa.eu/en/publication-detail/-/publication/78ae5276-ae8e-11e9-9d01-01aa75ed71a1/language-en

[70] European Commission. 2020. Commission report: EU data protection rules empower citizens and are fit for the digital age. https://ec.europa.eu/commission/presscorner/detail/en/ip_20_1163

[71] European Union. 1995. Directive of the European Parliament and the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data [1995] OJ L281/31. *Official Journal of the European Union* L281 (1995), 1–20. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31995L0046

[72] European Union. 2000. Charter of Fundamental Rights of the European Union. *OJ* 326 (2000), 391–407. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT

[73] European Union. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union* L119 (2016), 1–88. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC

[74] European Union. 2019. Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. *Official Journal of the European Union* L130 (2019), 92–125. https://eur-lex.europa.eu/eli/dir/2019/790/oj

[75] European Union. 2021. Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act). *Council of the European Union* 14606/21 (2021), 1–61. https://data.consilium.europa.eu/doc/document/ST-14606-2021-INIT/en/pdf

[76] Fair&Smart. 2016. Fair&Smart Personal Data Solutions. https://www.fairandsmart.com

[77] Todd Feathers. 2020. Colleges Say They Don't Need Exam Surveillance Tools to Stop Cheating. https://www.vice.com/en/article/88ag8z/colleges-say-they-dont-need-exam-surveillance-tools-to-stop-cheating

[78] Arthur Feinberg, Amineh Ghorbani, and Paulien Herder. 2021. Diversity and Challenges of the Urban Commons: A Comprehensive Review. *International Journal of the Commons* 15, 1 (Apr 2021), 1–20. https://doi.org/10.5334/ijc.1033

[79] Paul Feldman. 2020. Education and research improves lives, and technology improves education and research. https://www.foundation.org.uk/getattachment/8803ab67-86b4-4657-9dae-733a253e4741/paul-feldman-slides-pdf.pdf

[80] Tommaso Fia. 2021. An Alternative to Data Ownership: Managing Access to Non-Personal Data through the Commons. *Global Jurist* 21, 1 (2021), 181–210. https://doi.org/10.1515/gj-2020-0034

[81] Joshua B. Fisher and Louise Fortmann. 2010. Governing the data commons: Policy, practice, and the advancement of science. *Information & Management* 47, 4 (2010), 237–245. https://doi.org/10.1016/j.im.2010.04.001

[82] Colleen Flaherty. 2020. Big Proctor. https://www.insidehighered.com/news/2020/05/11/online-proctoring-surging-during-covid-19

[83] Andrea Forte, Vanesa Larco, and Amy Bruckman. 2009. Decentralization in Wikipedia Governance. *Journal of Management Information Systems* 26, 1 (2009), 49–72. https://doi.org/10.2753/MIS0742-1222260103

[84] Marianna Fotaki, Gaz Islam, and Anne Antoni. 2019. *Business Ethics and Care in Organizations*. Routledge, New York, USA.

[85] FreeYourMusic. 2021. FreeYourMusic homepage. https://freeyourmusic.com/.

[86] Batya Friedman, Peyina Lin, and Jessica Miller. 2005. Informed consent by design. *Security and Usability* 24, 1 (01 2005), 495–521.

[87] Archon Fung. 2015. Putting the Public Back into Governance: The Challenges of Citizen Participation and Its Future. *Public Administration Review* 75, 4 (2015), 513–522. https://doi.org/10.1111/puar.12361

[88] gE.CO. 2020. gE.CO Living Lab About Page. https://generative-commons.eu/

[89] Gabriel Geiger. 2021. Students Are Easily Cheating 'State-of-the-Art' Test Proctoring Tech. https://www.vice.com/en/article/3an98j/students-are-easily-cheating-state-of-the-art-test-proctoring-tech

[90] Samuel Gibbs. 2014. Twitter just made every public tweet findable ... here's how to delete yours. https://www.theguardian.com/technology/2014/nov/19/new-twitter-search-makes-every-public-tweet-since-2006-findable

[91] Mehitabel Glenhaber. 2020. A comprehensive guide to tech ethics and Zoom. https://sourceful.us/doc/652/a-comprehensive-guide-to-tech-ethics-and-zoom

[92] Bryan Glick. 2014. NHS England faces growing pressure to delay Care.data medical records plan. https://www.computerweekly.com/news/2240214577/NHS-England-faces-growing-pressure-to-delay-Caredata-medical-records-plan

[93] Ben Goldacre. 2014. Care.data is in chaos. It breaks my heart. https://www.theguardian.com/commentisfree/2014/feb/28/care-data-is-in-chaos

[94] Google. 2019. Street View. https://www.google.com/streetview/

[95] Neil Gordon. 2021. Flexible Pedagogies: technology-enhanced learning. https://www.advance-he.ac.uk/knowledge-hub/flexible-pedagogies-technology-enhanced-learning

[96] Don Gotterbarn, Amy Bruckman, Catherine Flick, Keith Miller, and Marty J. Wolf. 2018. ACM Code of Ethics: A Guide For Positive Action. https://cacm.acm.org/magazines/2018/1/223896-acm-code-of-ethics/fulltext

[97] Inge Graef, Martin Husovec, and Nadezhda Purtova. 2017. Data Portability and Data Control: Lessons for an Emerging Concept in EU Law. https://doi.org/10.2139/ssrn.3071875

[98] Maximilian Grafenstein. 2019. Co-Regulation and the Competitive Advantage in the GDPR: Data Protection Certification Mechanisms, Codes of Conduct and the 'State of the Art' of Data Protection-by-Design. *Forthcoming in González-Fuster, G., van Brakel, R. and P. De Hert Research Handbook on Privacy and Data Protection Law. Values, Norms and Global Politics* 1, 1 (18 Feb. 2019), 34. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3336990

[99] Yael Grauer. 2016. Dark Patterns are designed to trick you (and they're all over the Web). https://arstechnica.com/information-technology/2016/07/dark-patterns-are-designed-to-trick-you-and-theyre-all-over-the-web/

[100] Colin M. Gray, Cristiana Santos, Nataliia Bielova, Michael Toth, and Damian Clifford. 2021. *Dark Patterns and the Legal Requirements of Consent Banners: An Interaction Criticism Perspective*. Association for Computing Machinery, New York, USA, Chapter 1, 1–18. https://doi.org/10.1145/3411764.3445779

[101] Paul Grewal. 2018. Suspending Cambridge Analytica and SCL Group From Facebook. https://newsroom.fb.com/news/2018/03/suspending-cambridge-analytica/

[102] Robert L. Grossman, Allison Heath, Mark Murphy, Maria Patterson, and Walt Wells. 2016. A Case for Data Commons: Toward Data Science as a Service. *Computing in Science Engineering* 18, 5 (2016), 10–20. https://doi.org/10.1109/MCSE.2016.92

[103] Andrés Guadamuz. 2006. Open science: open source licences for scientific research. *North Carolina Journal of Law and Technology* 6 (2006), 321–366. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=886906

[104] National Data Guardian. 2016. National Data Guardian for Health and Care: Review of Data Security,Consent and Opt-Outs. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/535024/data-security-review.PDF

[105] Anita Gurumurthy and Nandini Chami. 2022. Governing the Resource of Data: To what end and for whom? https://datagovernance.org/report/governing-the-resource-of-data-to-what-end-and-for-whom-conceptual-building-blocks-of-a-semi-commons-approach.

[106] Jack Hardinges. 2020. Data trusts in 2020. https://theodi.org/article/data-trusts-in-2020/

[107] Rebecca Heilweil. 2020. Paranoia about cheating is making online education terrible for everyone. https://www.vox.com/recode/2020/5/4/21241062/schools-cheating-proctorio-artificial-intelligence

[108] Natali Helberger, Jo Pierson, and Thomas Poell. 2018. Governing online platforms: From contested to cooperative responsibility. *The Information Society* 34, 1 (2018), 1–14. https://doi.org/10.1080/01972243.2017.1391913

[109] Tristan Henderson. 2017. *Does the GDPR help or hinder fair algorithmic decision-making?* LLM dissertation. Edinburgh Law School, Edinburgh, UK. https://doi.org/10.2139/ssrn.3140887

[110] Alex Hern and David Pegg. 2018. Facebook fined for data breaches in Cambridge Analytica scandal. https://www.theguardian.com/technology/2018/jul/11/facebook-fined-for-data-breaches-in-cambridge-analytica-scandal

[111] Michael Herrmann, Mireille Hildebrandt, Laura Tielemans, and Claudia Diaz. 2016. Privacy in Location-Based Services: An Interdisciplinary Approach. *SCRIPTed* 13 (2016), 27. https://doi.org/10.2966/scrip.130216.144

[112] Charlotte Hess. 2006. Research on the Commons, Common-Pool Resources, and Common Property. https://dlc.dlib.indiana.edu/dlc/contentguidelines

[113] Charlotte Hess and Elinor Ostrom. 2007. *Understanding Knowledge as a Commons: From theory to practice*. MIT Press, Cambridge, USA.

[114] Mireille Hildebrandt and Laura Tielemans. 2013. Data protection by design and technology neutral law. *Computer Law & Security Review* 29, 5 (2013), 509–521. https://doi.org/10.1016/j.clsr.2013.07.004

[115] Kashmir Hill. 2012. How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did. https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/

[116] Kashmir Hill. 2018. The Secretive Company That Might End Privacy as We Know It. https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html

[117] Kashmir Hill and Surya Mattu. 2018. The House that Spied on Me. https://gizmodo.com/the-house-that-spied-on-me-1822429852

[118] Hong Kong Free Press. 2021. UK university tells lecturers not to record classes about Hong Kong and China, citing security law risks. https://hongkongfp.com/2021/05/10/uk-university-tells-lecturers-not-to-record-classes-about-hong-kong-and-china-citing-security-law-risks/

[119] Jane C. Hu. 2020. Paranoia about cheating is making online education terrible for everyone. https://slate.com/technology/2020/10/online-proctoring-proctoru-proctorio-cheating-research.html

[120] Mary Hui. 2019. Hong Kongers are using blockchain archives to fight government censorship. https://qz.com/2008673/hong-kongers-use-blockchain-to-fight-government-censorship/

[121] Human-Centred Artificial Intelligence, Stanford University. 2021. Radical Proposal: Data Cooperatives Could Give Us More Power Over Our Data. https://hai.stanford.edu/news/radical-proposal-data-cooperatives-could-give-us-more-power-over-our-data

[122] Human Data Interaction. 2018. Human Data Interaction and the HDI Network Plus. https://hdi-network.org

[123] IDA Ireland. 2019. Ireland's Corporate Tax Rate. https://www.idaireland.com/invest-in-ireland/ireland-corporate-tax

[124] Information Comissioner's Office. 2021. EU regulatory oversight. https://ico.org.uk/for-organisations/dp-at-the-end-of-the-transition-period/data-protection-and-the-eu-in-detail/the-uk-gdpr/eu-regulatory-oversight/

[125] International Co-operative Alliance. 2018. Cooperative identity, values & principles. https://www.ica.coop/en/cooperatives/cooperative-identity

[126] Involve. 2019. Designing decision making processes for data trusts: lessons from three pilots. https://www.involve.org.uk/sites/default/files/field/attachemnt/General-decision-making-report-Apr-19.pdf

[127] Ireland Data Protection Commission. 2018. One Stop Shop. https://www.dataprotection.ie/en/organisations/international-transfers/one-stop-shop-oss

[128] Chaminda Jayanetti. 2018. NHS data grab on hold as millions opt out. https://www.theguardian.com/society/2021/aug/22/nhs-data-grab-on-hold-as-millions-opt-out

[129] Mark A. Jensen, Vincent Ferretti, Robert L. Grossman, and Louis M. Staudt. 2017. The NCI Genomic Data Commons as an engine for precision medicine. *Blood* 130, 4 (2017), 453–459. https://doi.org/10.1182/blood-2017-03-735654

[130] JISC. 2020. Learning and teaching reimagined: a new dawn for higher education? https://www.jisc.ac.uk/reports/learning-and-teaching-reimagined-a-new-dawn-for-higher-education

[131] JISC. 2021. Learning analytics. https://www.jisc.ac.uk/learning-analytics

[132] Jumbo Privacy. 2019. Jumbo Privacy homepage. https://www.jumboprivacy.com/

[133] Dimitra Kamarinou, Christopher Millard, and W. Kuan Hon. 2016. Cloud privacy: an empirical study of 20 cloud providers' terms and privacy policies—Part I. *International Data Privacy Law* 6, 2 (06 2016), 79–101. https://doi.org/10.1093/idpl/ipw003

[134] Bert-Jaap Koops, Bryce Clayton Newell, Tjerk Timan, Ivan Škorvánek, Tom Chokrevski, and Maša Galič. 2016. A Typology of Privacy. *University of Pennsylvania Journal of International Law* 38 (2016), 483–575. https://ssrn.com/abstract=2754043

[135] Susanne P. Lajoie. 2005. Extending the Scaffolding Metaphor. *Instructional Science* 33, 5 (2005), 541–557. https://doi.org/10.1007/s11251-005-1279-2

[136] Lawrence Lessig. 2006. *Code: And Other Laws of Cyberspace, Version 2.0* (2 ed.). Basic Books, New York, USA.

[137] Peter Lewis. 2020. Peter Lewis's 2020s vision: stop glibly signing over your data and take control. https://www.theguardian.com/commentisfree/2020/feb/22/peter-lewiss-2020s-vision-stop-glibly-signing-over-your-data-and-take-control

[138] Wenlong Li. 2018. A tale of two rights: exploring the potential conflict between right to data portability and right to be forgotten under the General Data Protection Regulation. *International Data Privacy Law* 8, 4 (2018), 309–317. https://doi.org/10.1093/idpl/ipy007

[139] Andrew Liptak. 2018. Cambridge Analytica's use of Facebook data was a 'grossly unethical experiment'. https://www.theverge.com/2018/3/18/17134270/cambridge-analyticas-facebook-data-underscores-critical-flaw-american-electorate

[140] Natasha Lomas. 2021. Ireland's draft GDPR decision against Facebook branded a joke. https://techcrunch.com/2021/10/13/irelands-draft-gdpr-decision-against-facebook-branded-a-joke

[141] Samuel Macbeth. 2014. *Multi-agent based simulation of self-governing knowledge commons*. PhD Thesis. Imperial College London, London, UK. https://hdl.handle.net/10044/1/25751

[142] Callum MacDonald. 2007. Google's Street View site raises alarm over privacy. https://web.archive.org/web/20090213053639/http://www.theherald.co.uk/news/other/display.var.1444323.0.0.php

[143] Michael J. Madison. 2020. *Data governance and the emerging university*. Edward Elgar Publishing, Cheltenham, UK, Chapter 17, 364–390. https://www.elgaronline.com/view/edcoll/9781788116626/9781788116626.00027.xml

[144] Rene Mahieu, Hadi Asghari, and Michel van Eeten. 2017. Collectively Exercising the Right of Access: Individual Effort, Societal Effect. In *GigaNet (Global Internet Governance Academic Network) Annual Symposium 2017*. GigaNet, Geneva, Switzerland, 21. https://doi.org/10.2139/ssrn.3107292

[145] Shibani Mahtani. 2019. Large, peaceful protest shows Hong Kong's pro-democracy movement is still strong. https://www.washingtonpost.com/world/asia_pacific/half-a-year-on-hong-kongs-pro-democracy-movement-shows-its-still-strong/2019/12/08/f7314d92-17ee-11ea-80d6-d0ca7007273f_story.html

[146] Making Sense. 2014. Citizen sensing: a toolkit. https://making-sense.eu/wp-content/uploads/2018/01/Citizen-Sensing-A-Toolkit.pdf

[147] Naresh K. Malhotra, Sung S. Kim, and James Agarwal. 2004. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research* 15 (2004), 336–355. https://doi.org/10.1287/isre.1040.0032

[148] Edward J. Maloney and Joshua Kim. 2019. Intellectual Property and Digital Learning. https://www.insidehighered.com/digital-learning/blogs/technology-and-learning/intellectual-property-and-digital-learning

[149] Essam Mansour, Andrei Vlad Sambra, Sandro Hawke, Maged Zereba, Sarven Capadisli, Abdurrahman Ghanem, Ashraf Aboulnaga, and Tim Berners-Lee. 2016. A Demonstration of the Solid Platform for Social Web Applications. In *Proceedings of the 25th International Conference Companion on World Wide Web* (Montréal, Québec, Canada) *(WWW '16 Companion)*. International

World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 223–226. https://doi.org/10.1145/2872518.2890529

[150] Paul Mason. 2015. The Spotify privacy backlash: what is my personal data really worth? https://www.theguardian.com/commentisfree/2015/aug/23/the-spotify-privacy-backlash-what-is-my-personal-data-really-worth

[151] Sean McDonald. 2021. Data Governance's New Clothes. https://www.cigionline.org/articles/data-governances-new-clothes/

[152] Michael D. McGinnis. 2018. *The IAD Framework in Action: Understanding the Source of the Design Principles in Elinor Ostrom's Governing the Commons*. Lexington, Lanham, USA, 87–108. https://mcginnis.pages.iu.edu/McG-IAD%20applied%20book%20version.pdf

[153] Timothy McLaughlin. 2021. The end of free speech in Hong Kong. https://www.theatlantic.com/international/archive/2021/07/end-free-speech-hong-kong/619577/

[154] Sandra J. Milberg, Sandra J. Burke, H. Jeff Smith, and Ernest A. Kallman. 1995. Values, Personal Information Privacy, and Regulatory Approaches. *Commun. ACM* 38, 12 (Dec. 1995), 65–74. https://doi.org/10.1145/219663.219683

[155] Darakhshan J. Mir. 2021. *Designing for the Privacy Commons*. Cambridge University Press, Cambridge, UK, 245–267. https://doi.org/10.1017/9781108749978.011

[156] Danny Mok, William Yiu, Emily Tsang, and Jack Tsang. 2021. University of Hong Kong removes Pillar of Shame sculpture marking Tiananmen Square crackdown in middle of night. https://www.scmp.com/news/hong-kong/article/3160744/university-hong-kong-covers-pillar-shame-sculpture-marking-tiananmen

[157] Mayo Fuster Morell, Jorge L Salcedo, and Marco Berlinguer. 2016. Debate About the Concept of Value in Commons-Based Peer Production. *International Conference on Internet Science* 9934, 1 (Aug 2016), 27–41. https://doi.org/10.1007/978-3-319-45982-0_3

[158] Ekaterina Muravyeva, José Janssen, Marcus Specht, and Bart Custers. 2020. Exploring solutions to the privacy paradox in the context of e-assessment: informed consent revisited. *Ethics and Information Technology* 22, 1 (04 2020), 223–238. https://doi.org/10.1007/s10676-020-09531-5

[159] National Cancer Institute. 2020. Genomic Data Commons. https://gdc.cancer.gov/

[160] New Economics Foundation. 2018. Co-operatives unleashed: Doubling the size of the UK's co-operative sector. https://neweconomics.org/uploads/files/co-ops-unleashed.pdf

[161] Newspeak House. 2020. Coronavirus Tech Handbook. https://coronavirustechhandbook.com/

[162] Facebook Newsroom. 2018. Shutting Down Partner Categories. https://newsroom.fb.com/news/h/shutting-down-partner-categories/

[163] Davide Nicolini, Jeanne Mengis, and Jacky Swan. 2012. Understanding the Role of Objects in Cross-Disciplinary Collaboration. *Organization Science* 23, 3 (2012), 612–629. https://doi.org/10.1287/orsc.1110.0664

[164] Helen Nissenbaum. 2004. Privacy as Contextual Integrity. *Washington Law Review* 79 (2004), 119–158. https://heinonline.org/HOL/Page?handle=hein.journals/washlr79&id=129&collection=journals&index=

[165] Clive Norris, Paul de Hert, Xavier L'Hoiry, and Antonella Galetta. 2017. *The Unaccountable State of Surveillance*. Vol. 34. Springer International Publishing, Cham, Switzerland. https://doi.org/10.1007/978-3-319-47573-8

[166] Midas Nouwens, Ilaria Liccardi, Michael Veale, David Karger, and Lalana Kagal. 2020. Dark Patterns after the GDPR: Scraping Consent Pop-Ups and Demonstrating Their Influence. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, USA, 1–13. https://doi.org/10.1145/3313831.3376321

[167] Office for Students. 2021. Gravity assist: propelling higher education towards a brighter future. https://www.officeforstudents.org.uk/publications/gravity-assist-propelling-higher-education-towards-a-brighter-future/executive-summary/

[168] Kieron O'Hara. 2019. Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship. https://eprints.soton.ac.uk/428276/1/WSI_White_Paper_1.pdf

[169] Open Data Institute. 2019. Data trusts: lessons from three pilots. https://docs.google.com/document/d/118RqyUAWP3WIyyCO4iLUT3oOobnYJGibEhspr2v87jg/edit

[170] Open Data Institute. 2021. ODI Fellow Report: Data governance for online learning. https://theodi.org/article/data-governance-online-learning/

[171] Open Data Institute. 2021. What are data institutions and why are they important? https://theodi.org/article/what-are-data-institutions-and-why-are-they-important/

[172] Open Usage. 2021. Open Usage Commons. https://openusage.org/

[173] OpenGDPR. 2018. OpenGDPR. https://github.com/opengdpr/opengdpr

[174] Elinor Ostrom. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge University Press, Cambridge, UK.

[175] Elinor Ostrom. 2005. *Understanding Institutional Diversity* (1 ed.). Princeton University Press, Princeton, USA.

[176] Elinor Ostrom. 2010. Polycentric systems for coping with collective action and global environmental change. *Global Environmental Change* 20 (2010), 550–557. https://doi.org/10.1016/j.gloenvcha.2010.07.004

[177] Elinor Ostrom. 2012. *The Future of the Commons: Beyond Market Failure & Government Regulations*. Institute of Economic Affairs, London, UK.

[178] Vincent Ostrom, Charles M. Tiebout, and Robert Warren. 1961. The organization of government in metropolitan areas: a theoretical inquiry. *American Political Science Review* 55 (1961), 831–842. https://doi.org/10.1017/S0003055400125973

[179] P2P Foundation Wiki. 2021. Data Cooperatives. https://wiki.p2pfoundation.net/Data_Cooperatives

[180] Yuliya Panfil and Andrew Hagopian. 2019. A Commons Approach to Data Governance. https://www.newamerica.org/weekly/commons-approach-to-data-governance/.

[181] Shayna Pekala. 2017. Privacy and User Experience in 21st Century Library Discovery. *Information Technology and Libraries* 36, 2 (Jun. 2017), 48–58. https://doi.org/10.6017/ital.v36i2.9817

[182] Alex Pentland. 2013. The Data-Driven Society. *Scientific American* 309 (10 2013), 78–83. https://www.scientificamerican.com/article/how-big-data-can-transform-society-for-the-better/

[183] Aidan Peppin. 2020. Doing good with data: what does good look like when it comes to data stewardship? https://www.adalovelaceinstitute.org/doing-good-with-data-what-does-good-look-like-when-it-comes-to-data-stewardship/

[184] Jeremy Pitt. 2012. Managing User-Generated Content as a Knowledge Commons. *Logic Programs, Norms and Action* 7360, 1 (Jan 2012), 401–424. https://doi.org/10.1007/978-3-642-29414-3_21

[185] Jeremy Pitt and Julia Schaumeier. 2012. Provision and Appropriation of Common-Pool Resources without Full Disclosure. *PRIMA 2012: Principles and Practice of Multi-Agent Systems* 7455, 1 (Sep 2012), 199–213. https://doi.org/10.1007/978-3-642-32729-2_14

[186] Leah Plunkett, Urs Gasser, and Sandra Cortesi. 2021. Student Privacy and the Law in the Internet Age. *The Oxford Handbook of U.S. Education Law* 1, 1 (2021), 24. https://doi.org/10.1093/oxfordhb/9780190697402.013.30

[187] Port.im. 2015. Port.im homepage. https://www.port.im/

[188] Alison B. Powell. 2021. Explanations as governance? Investigating practices of explanation in algorithmic system design. *European Journal of Communication* 36, 4 (2021), 362–375. https://doi.org/10.1177/02673231211028376

[189] Alison B. Powell. 2021. Open culture and innovation: integrating knowledge across boundaries. *Media, Culture and Society* 37 (2021), 376–393. https://doi.org/10.1177/0163443714567169

[190] Alison B. Powell. 2021. *Undoing Optimization*. Yale University Press, New Haven, USA.

[191] Michael Powell. 2021. What's the difference between a foundation and a trust? https://www.hawksford.com/knowledge-hub/2014/foundations-vs-trusts

[192] Barbara Prainsack. 2019. Logged out: Ownership, exclusion and public value in the digital data and information commons. *Big Data & Society* 6, 1 (2019), 1–15. https://doi.org/10.1177/2053951719829773

[193] Paul Prinsloo and Sharon Slade. 2016. Student vulnerability, agency, and learning analytics: An exploration. *Journal of Learning Analytics* 3 (2016), 159–182. https://doi.org/10.18608/jla.2016.31.10

[194] Paul Prinsloo and Sharon Slade. 2018. Student Consent in Learning Analytics: The Devil in the Details? In *Learning Analytics in Higher Education: Current Innovations, Future Potential, and Practical Applications*, Jaime Lester, Carrie Klein, Aditya Johri, and Huzefa Rangwala (Eds.). Routledge, New York and Abingdon, Oxon, 118–139. https://oro.open.ac.uk/55361/

[195] Privacy International. 2021. Ad-tech. https://privacyinternational.org/learn/adtech

[196] Peter Rabley and Christopher Keefe. 2021. Establishing a Data Trust: It's Really Hard. https://www.thisisplace.org/blog-1/introducingplace/its-really-hard

[197] Joseph Raz. 1986. *The Morality of Freedom* (1 ed.). Oxford University Press, Oxford, UK. 208 pages.

[198] Chris Reed and Murray Andrews. 2018. *Rethinking the Jurisprudence of Cyberspace* (1 ed.). Elgar Publishing, Cheltenham, UK.

[199] Beate Roessler. 2005. *The Value of Privacy* (1 ed.). Polity, Cambridge, UK. 51 pages.

[200] Neil Rose. 2020. Bar students urge online exams rethink. https://www.legalfutures.co.uk/latest-news/bar-students-urge-online-exams-rethink

[201] Eleanor Ainge Roy. 2018. Facebook data breach hits 63,714 New Zealanders after 10 people download quiz. https://www.theguardian.com/technology/2018/apr/10/facebook-data-breach-hits-63714-new-zealanders-after-10-people-download-quiz

[202] Paul Ryan, Martin Crane, and Rob Brennan. 2021. GDPR Compliance Tools: Best Practice from RegTech. *ICEIS 2020: Enterprise Information Systems* 417, 1 (2021), 905–929. https://doi.org/10.1007/978-3-030-75418-1_41

[203] William A. Sandoval and Brian J. Reiser. 2004. Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education* 88, 3 (2004), 345–372. https://doi.org/10.1002/sce.10130

[204] Madelyn Sanfilippo, Brett Frischmann, and Katherine Standburg. 2018. Privacy as Commons: Case Evaluation Through the Governing Knowledge Commons Framework. *Journal of Information Policy* 8 (2018), 116–166. https://www.jstor.org/stable/10.5325/jinfopoli.8.2018.0116

[205] Susanna-Assunta Sansone, Peter McQuilton, Philippe Rocca-Serra, Alejandra Gonzalez-Beltran, Massimiliano Izzo, Allyson L. Lister, Milo Thurston, and the FAIRsharing Community. 2019. FAIRsharing as a community approach to standards, repositories and policies. *Nature Biotechnology* 37 (2019), 358–367. https://doi.org/10.1038/s41587-019-0080-8

[206] Bruce Schneier. 2006. The eternal value of privacy. https://www.wired.com/2006/05/the-eternal-value-of-privacy/

[207] m.c. schraefel, Richard Gomer, Alper Alan, Enrico Gerding, and Carsten Maple. 2017. The Internet of Things: Interaction Challenges to Meaningful Consent at Scale. *Interactions* 24, 6 (Oct. 2017), 26–33. https://doi.org/10.1145/3149025

[208] Mike Schroepfer. 2018. An Update on Our Plans to Restrict Data Access on Facebook. https://newsroom.fb.com/news/2018/04/restricting-data-access/

[209] Driver's Seat. 2020. Driver's Seat homepage. https://driversseat.co/

[210] Irina Shklovski. 2018. Privacy as ability or a state: An argument for a relational view. https://blogit.itu.dk/virteuproject/2019/10/28/

privacy-as-ability-or-a-state-an-argument-for-a-relational-view/

[211] Sarah Silverman, Autumm Caines, Christopher Casey, Belen Garcia de Hurtado, Jessica Riviere, Alfonso Sintjago, and Carla Vecchiola. 2021. What Happens When You Close the Door on Remote Proctoring? Moving Toward Authentic Assessments with a People-Centered Approach. *Educational Development in the Time of Crises* 39, 3 (2021), 18. https://doi.org/10.3998/tia.17063888.0039.308

[212] Parmider Jeet Singh and Jai Vipra. 2019. Economic Rights Over Data: A Framework for Community Data Ownership. *Development* 62 (2019), 53–57.

[213] Daniel J. Solove. 2010. *Understanding Privacy* (1 ed.). Harvard University Press, Cambridge, USA.

[214] Jared Spataro. 2020. Our commitment to privacy in Microsoft Productivity Score. https://www.microsoft.com/en-us/microsoft-365/blog/2020/12/01/our-commitment-to-privacy-in-microsoft-productivity-score/

[215] Felix Stalder. 2010. Digital commons. In *The human economy. A citizen's guide*, Keith Hart, Jean-Louis Laville, and Antonio David Cattani (Eds.). Polity Press, Cambridge, UK, 313–324.

[216] Sophie Stalla-Bourdillon, Laura Carmichael, and Alexsis Wintour. 2021. Fostering trustworthy data sharing: Establishing data foundations in practice. *Data & Policy* 3 (2021), e4. https://doi.org/10.1017/dap.2020.24

[217] Sophie Stalla-Bourdillon, Gefion Thuermer, Johanna Walker, Laura Carmichael, and Elena Simperl. 2020. Data protection by design: Building the foundations of trustworthy data sharing. *Data & Policy* 2 (2020), e4. https://doi.org/10.1017/dap.2020.1

[218] Sophie Stalla-Bourdillon, Alexsis Wintour, and Laura Carmichael. 2021. Building Trust Through Data Foundations; A Call for a Data Governance Model to Support Trustworthy Data Sharing. https://eprints.soton.ac.uk/443715/

[219] Paul C. Stern. 2011. Design principles for global commons: natural resources and emerging technologies. *International Journal of the Commons* 5, 2 (2011), 213–232. https://doi.org/10.18352/ijc.305/

[220] Jack Stilgoe, Simon J. Lock, and James Wilsdon. 2014. Why should we promote public engagement with science? *Public Understanding of Science* 23, 1 (2014), 4–15. https://doi.org/10.1177/0963662513518154 PMID: 24434705.

[221] Chris Stokel-Walker. 2020. Universities are using surveillance software to spy on students. https://www.wired.co.uk/article/university-covid-learning-student-monitoring

[222] Katherine J. Strandburg, Brett M. Frischmann, and Michael J. Madison. 2017. *The Knowledge Commons Framework*. Cambridge University Press, Cambridge, UK, 9–18. https://doi.org/10.1017/9781316544587.002

[223] Joanna Strycharz, Jef Ausloos, and Natali Helberger. 2020. Data Protection or Data Frustration? Individual Perceptions and Attitudes Towards the GDPR. *European Data Protection Law Review* 6, 3 (2020), 407–42. https://doi.org/10.21552/edpl/2020/3/10

[224] Iryna Susha, Marijn Janssen, and Stefaan Verhulst. 2017. Data collaboratives as "bazaars"? A review of coordination problems and mechanisms to match demand for data with supply. *Transforming Government: People, Process and Policy* 11 (2017), 24. https://doi.org/10.1108/TG-01-2017-0007

[225] Shea Swauger. 2020. Our Bodies Encoded: Algorithmic Test Proctoring in Higher Education. https://hybridpedagogy.org/our-bodies-encoded-algorithmic-test-proctoring-in-higher-education/

[226] Peter Swire and Yianni Lagos. 2013. Why the Right to Data Portability Likely Reduces Consumer Welfare: Antitrust and Privacy Critique. *Maryland Law Review* 72, 2 (2013), 335–380. https://doi.org/10.2139/ssrn.2159157

[227] Charlie Taylor. 2018. Data Protection Commission 'disappointed' at budget allocation. https://www.irishtimes.com/business/technology/data-protection-commission-disappointed-at-budget-allocation-1.4045248

[228] Charlie Taylor. 2022. Emily O'Reilly opens inquiry into European Commission policing of GDPR in Ireland. https://www.irishtimes.com/business/technology/emily-o-reilly-opens-inquiry-into-european-commission-policing-of-gdpr-in-ireland-1.4798907

[229] Linnet Taylor. 2017. Conclusion: What Do We Know About Group Privacy? In *Group Privacy: New Challenges of Data Technologies"*, Linnet Taylor, Luciano Floridi, and Bart van der Sloot (Eds.). Springer International Publishing, London, UK, 225–237. https://doi.org/10.1007/978-3-319-46608-8_12

[230] The Bristol Approach. 2017. The Bristol Approach homepage. https://www.bristolapproach.org/

[231] The Data Transfer Project. 2018. Data Transfer Project homepage. https://datatransferproject.dev/

[232] The Global Partnership of Artificial Intelligence. 2021. Enabling data sharing for social benefit through data trusts. https://gpai.ai/projects/data-governance/data-trusts/

[233] The Governance Lab. 2021. Data Collaboratives. https://datacollaboratives.org/

[234] The Governance Lab. 2021. Wanted: Data Stewards: (Re-)Defining The Roles and Responsibilities of Data Stewards for an Age of Data Collaboration. https://www.thegovlab.org/static/files/publications/wanted-data-stewards.pdf

[235] The GPAI Data Governance Working Group. 2021. Understanding Data Trusts. https://ceimia.org/wp-content/uploads/2021/07/2021-07-09-GPAI-summary-understanding-data-trusts-updated.docx.pdf

[236] The Royal Society. 2018. Protecting privacy in practice: The current use, development and limits of Privacy Enhancing Technologies in data analysis. https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/privacy-enhancing-technologies-report.pdf

[237] The Scotsman. 2018. Edinburgh University lecture recordings used against strikes. https://www.scotsman.com/education/edinburgh-university-lecture-recordings-used-against-strikes-332569

[238] Dennis F. Thompson. 1980. Moral responsibility of public officials: The problem of many hands. *The American Political Science Review* 74, 5 (1980), 905–16. https://doi.org/10.2307/1954312

[239] Dennis F. Thompson. 2014. Responsibility for Failures of Government: The Problem of Many Hands. *The American Review of Public Administration* 44, 3 (2014), 259–273. https://doi.org/10.1177/0275074014524013

[240] Manuel Tironi. 2015. Disastrous Publics: Counter-enactments in Participatory Experiments. *Science, Technology, & Human Values* 40, 4 (2015), 564–587. https://doi.org/10.1177/0162243914560649

[241] Rebekah Tromble and Daniela Stockmann. 2017. *Lost Umbrellas: Bias and the Right to Be Forgotten in Social Media Research*. Peter Lang, Oxford, UK, Chapter 5, 75–90. https://doi.org/10.3726/b11077

[242] Zeynep Tufecki. 2017. *Twitter and tear gas : the power and fragility of networked protest*. Yale University Press, New Haven, USA.

[243] Zeynep Tufecki. 2020. The Pandemic Is No Excuse to Surveil Students. https://www.theatlantic.com/technology/archive/2020/09/pandemic-no-excuse-colleges-surveil-students/616015/

[244] United Kingdom. 2018. Data Protection Act. *Act of Parliament* 1 (2018), 1–335. https://www.legislation.gov.uk/ukpga/2018/12/enacted/data.pdf

[245] United Kingdom. 2020. Freedom of Information Act. *Act of Parliament* 1 (2020), 1–173. https://www.legislation.gov.uk/ukpga/2000/36/data.pdf

[246] u/randomwordbot. 2019. Reddit post: Upset about Blizzard's HK ruling? Here's what to do about it. https://www.reddit.com/r/hearthstone/comments/df0zx5/upset_about_blizzards_hk_ruling_heres_what_to_do/. Accessed: 2022-01-18.

[247] Funda Ustek-Spilda, Alison B. Powell, and Selena Nemorin. 2019. Engaging with ethics in Internet of Things: Imaginaries in the social milieu of technology developers. *Big Data & Society* 6, 2 (2019), 2053951719879468. https://doi.org/10.1177/2053951719879468

[248] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019. (Un)Informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (London, United Kingdom) *(CCS '19)*. Association for Computing Machinery, New York, USA, 973–990. https://doi.org/10.1145/3319535.3354212

[249] Ibo van de Poel, Jessica Nihlén Fahlquist, Neelke Doorn, Sjoerd Zwart, and Lambèr Royakkers. 2012. The Problem of Many Hands: Climate Change as an Example. *Science and Engineering Ethics* 18, 1 (2012), 49–67. https://doi.org/10.1007/s11948-011-9276-0

[250] Frank Van Laerhoven and Clare Barnes. 2014. Communities and commons: the role of community development support in sustaining the commons. *Community Development Journal* 49, suppl_1 (01 2014), i118–i132. https://doi.org/10.1093/cdj/bsu005

[251] Michael Veale, Reuben Binns, and Jef Ausloos. 2018. When data protection by design and data subject rights clash. *International Data Privacy Law* 8, 2 (04 April 2018), 105–123. https://doi.org/10.1093/idpl/ipy002

[252] Stefaan G. Verhulst. 2021. Reimagining data responsibility: 10 new approaches toward a culture of trust in re-using data to address critical public needs. *Data & Policy* 3 (2021), 6. https://doi.org/10.1017/dap.2021.4

[253] Stefaan G. Verhulst and David Sangokoya. 2015. Data Collaboratives: Exchanging Data to Improve People's Lives. https://sverhulst.medium.com/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a

[254] Stefaan G. Verhulst, Andrew Young, and Prianka Srinivasan. 2021. An Introduction to Data Collaboratives: Creating Public Value by Exchanging Data. https://datacollaboratives.org/static/files/data-collaboratives-intro.pdf

[255] James Vincent. 2020. University staff are worried their recorded lectures will be used against them. https://www.theverge.com/21373669/recorded-lecture-capture-copyright-universities-coronavirus-fears

[256] Riina Vuorikari, Yves Punie, and Marcelino Cabrera Giraldez. 2020. Emerging technologies and the teaching profession. *J RC Science for Policy* 30129, 1 (2020), 68. https://doi.org/10.2760/46933

[257] Derek Wall. 2017. *Elinor Ostrom's Rules for Radicals* (1 ed.). Pluto Press, London, UK.

[258] Samuel Warren and Louis Brandeis. 1890. The Right to Privacy. *Harvard Law Review* 4 (1890), 193–220.

[259] Olivia B Waxman. 2018. The GDPR Is Just the Latest Example of Europe's Caution on Privacy Rights. That Outlook Has a Disturbing History. https://time.com/5290043/nazi-history-eu-data-privacy-gdpr/

[260] Stefan Weiss. 2009. Privacy threat model for data portability in social network applications. *International Journal of Information Management* 29, 4 (2009), 249–254. https://doi.org/10.1016/j.ijinfomgt.2009.03.007

[261] Katrin Weller and Katharina Kinder-Kurlanda. 2017. *To Share or Not to Share? Ethical Challenges in Sharing Social Media-based Research Data*. Peter Lang, Oxford, UK, Chapter 7, 115–129. https://doi.org/10.3726/b11077

[262] Alan F. Westin. 1967. *Privacy and Freedom* (1 ed.). Scribner, New York, USA.

[263] Wikimedia. 2021. Commons Project Scope. https://commons.wikimedia.org/wiki/Commons:Project_scope

[264] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffery S. Grethe, Jaap Heringa, Peter A.C't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Alber Mons, Abel L. Packer,

Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 1 (2016), 1–9. https://doi.org/10.1038/sdata.2016.18

[265] Ben Williamson, Rebecca Eynon, and John Potter. 2020. Pandemic politics, pedagogies and practices: digital technologies and distance education during the coronavirus emergency. *Learning, Media and Technology* 45, 2 (2020), 107–114. https://doi.org/10.1080/17439884.2020.1761641

[266] James Wilsdon and Rebecca Willis. 2004. *See-through science: Why public engagement needs to move upstream*. Demos, London, UK.

[267] James Wilsdon, Brian Wynne, and Jack Stilgoe. 2005. The public value of science. *London: Demos* 1 (2005), 69. https://sciencescitoyennes.org/wp-content/uploads/archives_doc/pdf/publicvalueofscience.pdf

[268] Janis Wong and Tristan Henderson. 2019. The right to data portability in practice: exploring the implications of the technologically neutral GDPR. *International Data Privacy Law* 9, 3 (07 2019), 173–191. https://doi.org/10.1093/idpl/ipz008

[269] David Wood, Jerome S. Bruner, and Gail Ross. 1976. The role of tutoring in problem-solving. *Journal of Child Psychology and Child Psychiatry* 17, 2 (1976), 89–100. https://doi.org/10.1111/j.1469-7610.1976.tb00381.x

[270] Nicola Woolcock. 2022. Lecturers admit self-censoring classes with Chinese students. https://www.thetimes.co.uk/article/lecturers-admit-self-censoring-classes-with-chinese-students-wjlf07lng

[271] Brian Wynne. 2006. Public Engagement as a Means of Restoring Public Trust in Science – Hitting the Notes, but Missing the Music? *Community Genetics* 9, 3 (2006), 211–220. https://www.jstor.org/stable/26679532

[272] Andrew Young and Stefaan G. Verhulst. 2020. *Data Collaboratives*. Springer International Publishing, Cham, Switzerland, 1–5. https://doi.org/10.1007/978-3-030-13895-0_92-1

[273] Shoshana Zuboff. 2015. Big other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology* 30, 1 (2015), 75–89. https://doi.org/10.1057/jit.2015.5

[274] Mark Zuckerberg. 2018. Cambridge Analytica Facebook Post. https://www.facebook.com/zuck/posts/10104712037900071

PART I

APPENDIX

APPENDIX A

INTERVIEW QUESTIONS

Interview Structure and Questions

Introduction

- Purpose of the interview:
  - Identifying and understanding challenges in the commons development process from the perspective of the interviewee.
- Participation:
  - Interviews for this project forms part of my doctoral research on developing a socio-technical "data commons" framework to better protect data subject's personal data.
  - As interviewees have been part of the process for developing a commons and have both experience and knowledge to inform how a commons can be successful.
- Interviews with be recorded electronically if the interviewee consents.

Establish common understanding of:

- What a commons is and why the PROJECT is considered a commons.
- What the interviewee's role is as part of the PROJECT.
- A summary of what the PROJECT is.

Key Questions 1: Project aims and identifying problems

- Can you tell me more about your role in the PROJECT and the PROJECT's aims?
- How did you and the PROJECT team come about identifying those aims and what were some of the problems or challenges you considered during that process?
  - How does/did your role fits into achieving those aims?
- What stakeholders did you interact with?
  - Was it useful? Why?
- What private actors and companies (e.g. digital service providers) did you work with?
  - What role did they play in the process of developing the project aims?
  - What is your understanding of the tools and services they provide?
  - How transparent were they?
  - What did you agree on in terms of the PROJECT?
  - Where there any disagreements? How were these resolved?
- How did the private actor (positively or negatively) influence the design, development, and architecture of the commons for your PROJECT?

Key Questions 2: Problem solving and overcoming challenges

- Were there problems or challenges during the PROJECT that you didn't expect?
  - How did you go about solving them?
- IF NONE WERE IDENTIFIED: Were problems or challenges related to data protection and privacy considered during the process?
  - Were they considered challenges?

- IF NONE WERE IDENTIFIED: Were problems or challenges related to the implementation of "smart" or IoT technologies considered during the process?
  - Were they considered challenges?

Key Questions 3: Overall project perspectives

- What do you think are/were the successes of your role in the PROJECT?
  - How did you go about achieving this?
- What do you think are/were the limitations of your role in the PROJECT?
  - Why?
- Do you think the PROJECT as a commons is/was successful?
  - Why?
- What do you think are/were the limitations of the PROJECT as a commons? What would you do differently?

Broader Questions

- What do you think is the ideal commons?
  - What are some theories or practices that you know of that has helped form your understanding of this ideal commons?
  - What do you think an ideal commons should achieve?
  - What stakeholders do you think should be involved?
    - How involved should data subjects and participants of the commons be in its creation?
    - Do you think anyone should be excluded from the process?
  - Is there anything you would like to add?

Summary

- Discussed your role, identifying problems and challenges from the commons process.
- Wrap up interview.
- Provide more details about our research and how this interview fits in.
- Stop recording for final summary and discussion to end the session.
- Provide another opportunity to ask questions or additional information.
- Timeline of our research.

Appendix B

Adapting and applying the data protection IAD commons framework

To create a data protection-focused data commons for online learning, we applied the data protection IAD framework for the use case of supporting students as data subjects in deciding and expressing their data protection preferences for online learning.

The questions that were identified as part of the data protection IAD framework are answered as follows:

Background

- The background context of the data protection-focused data commons for online learning involves the environment in which online learning is being undertaken and the requirement for tutorial recordings of classes. This was heightened by the COVID-19 pandemic that shifted all learning online.

  - As online learning progressed, more awareness came to light about the monitoring of students through technologies. Pre-pandemic, there were also considerations about the impact of new technologies, tutorial and lecture recordings, and the digitisation of education more generally.
  - Despite positive progress in containing the pandemic, institutions are continuing to adopt some of these technological practices even as in-person teaching is able to resume. As a result, it is important to ensure that students are able to understand how their data is used and have the ability to control that data.

- As part of existing regulations such as the GDPR, universities and HEIs have the responsibility to clarify and explain how they use personal data. Currently, universities have privacy policies on online learning and tutorial recordings as well as wider data protection impact assessments and policies. Universities also have DPOs as required by organisations of a certain size to respond to any data protection requests and answer any data protection-related issues.

  - Data protection-specific and sector-wide organisations such as the ICO, JISC, and the OfS outline what and how data should and should not be used in relation to the work environment and specifically for higher education. Some of this work pre-dates the COVID-19 pandemic. Some research was published during the pandemic in producing solutions that support the protection of personal data for the future of education.
  - Students' personal data is separate from other forms of data that universities manage. For example, students' administrative data, examination and assessments data, and data from tutorials are all managed differently by different departments within the university. However, there may be some overlap, highlighting the importance of students being able to control and understand how their personal data from tutorials is being processed. Universities generally follow FAIR data principles with regards to research data.
  - Students may not be aware of how their institution manages their data and may not feel like they can challenge their institution given that doing so could negatively impact both their academic experience as well as their grades.

- Trust issues between students and their institutions, as well as staff and their institutions, may have arisen based on incidences of technology adoption as well as the sharing of recordings without explicit consent.

Data Attributes

- The data and personal data that are part of the commons.
  - Student's personal data as part of Microsoft Teams such as student ID, the content they reveal in the tutorial, chat data, screen sharing, and their voice.
  - If they disclose any disabilities, racial information, religious information, political identities, or union membership, this could be classified as sensitive personal data.
  - The data is collected and processed following the university's policies, through Microsoft Teams, and possibly internationally if the student is not based within the UK.
  - University and third-party software collect, store, and process the data.
  - The data is stored privately although tutorial recordings may be shared with other students. Currently, students have limited control as to whether they want to be recorded.
  - University tutors, IT teams, and systems teams are responsible for how the data is stored, shared and retained, with different administrative privileges.
  - The university uses third-party software such as Microsoft Teams as well as Panopto to record and store recordings.
  - Students have limited control and authority in the process. They only have information of the university policies.
  - Some of the risks include extensive data gathering unrelated to education, potential discrimination from e-proctoring software, and creating a surveillance academic environment.

Commons Community Members

- The commons aims to support students and will also include staff, IT admin, and potential experts or those who are able to provide external advice outside

of the university.

- The commons is only relevant for those within the university community given that the data only applies to online learning.
- The technology companies that provide the tools for online learning as well as higher education organisations such as JISC or the OfS may be relevant for the commons.
- Students have a power imbalance between themselves, staff, university management, and potential employers given that if they refuse certain personal data to be collected or provided, they may not be able to access education, negatively impacting their academic prospects.

Goals and Objectives

- The objective of the commons is to support students' online learning personal data preferences and help them understand what data protection rights and recourse they have should they not want their personal data to be used in certain ways.

Managing and Governing the Commons

- The commons will sit on top of the online learning platform, in this case Microsoft Teams, to allow seamless and integrated access to the tool while not compromising students' privacy with respect to others in the tutorial.
  - The commons will allow students to choose whether they want to consent to tutorial recording both before and after the tutorial, with respect to the collection and processing of their personal data in that way.
  - Online learning data that is collected is shared within the tutorial and possibly to other students as well, where the recording may be re-purposed for teaching beyond the session in which the student participated in.
  - No data protection mechanisms currently exist for this use case and only university policies are applied.

  - The relevant data subject rights include the right of access, the right to data portability, and the right to object to automated decision-making.
  - Purpose limitation may have been considered, but is inconclusive.

- Determine the governance mechanisms of the commons.
  - The commons community consists of those who are affiliated with the university.
  - There is no requirement for those who participate to share their personal data or their experience, but in order for the commons to function and meet its aims, students need to vote as to whether they consent to tutorial recording.
  - If appropriate, the tutor can mitigate any issues. If not, then an external, neutral expert can help as well as addressing the DPO.
  - Existing platforms that are used to conduct online learning may be updated with better privacy support or offer tools that can better protection users' personal data.

- Identifying decision-makers and experts.
  - External experts can be identified to support the commons, such as academics from other institutions, privacy professionals, and independent or international higher education bodies.

- Decision-making on the commons is determined in part by the tutor, the department, and university management, with the latter making the most impact.
  - The commons would be digital and take place on the same platform as where the online learning is taking place.
  - Some of the infrastructure is internal, for example where the recording may be embedded and uploaded. Some of the infrastructure is external and provided by third-party companies.

- Establishing formal or informal norms that govern the commons.
  - The commons follows the same guidelines as the terms of service of the online learning provider as well as university policies.

  - Students and the commons community can provide feedback on their online learning experience through standard university procedures.
  - Some institutions, such as the Open University, have more experience with delivering online learning.

Outcomes

- Benefits of the commons.
  - Students are able to understand and control how their personal data is being used as well as what avenues there are to object against some uses of personal data.
  - The commons community should expect advice and guidance on what is allowed, as well as have the ability to anonymously share their experience with others.
- Costs and risks of the commons.
  - The commons has minimum risk given that no extra personal data is being collected. There are mechanisms in place to ensure that their consent vote is anonymous and cannot be traced back to them or the tutor. There are no risks of further data breaches or privacy problems.
  - As the tool is developed on Microsoft Teams and hosted by internal university servers, there are no additional risks from the data infrastructure.
  - The rights available under the GDPR apply to the commons where applicable to personal data.

Appendix C

Mock-tutorial Document

As part of the Online Learning as a Commons study, we would like you to imagine that you are participating in a Microsoft Teams-based tutorial. If you have not yet received a Microsoft Teams meeting invitation, please e-mail the researchers.

Please read the tutorial scenario below. Note that no further preparation will be needed before the Microsoft Teams meeting.

Tutorial: An Introduction to Conducting Research on Social Media

In our digitally connected society, social media such as Facebook, Twitter, LinkedIn, and Instagram are used not only for sharing parts of our lives with others, but also used by businesses, event organisers, recruiters, and data brokers to better understand how individuals and groups interact.

In this introduction, we will explore the types of data that are collected through social media, different techniques for conducting social media research, and review some examples and case studies.

This tutorial is aimed at a general audience and is suitable for all disciplines.

1. What is social media research?

Social media research is where quantitative or qualitative data is being gathered from social networking sites (SNS). This research can be done in many forms. Examples of social media research include:

- Downloading tweets from the Twitter Archive and looking at specific hashtags.
- Looking at the user engagement (such as views, clicks, and location) of an advertisement put out by a business on Facebook.
- Creating polls on Instagram and asking users specific questions.

Social media research can be conducted by individuals and businesses to improve their understanding of specific demographics of users to serve them specific content or find out more about their behaviours.

Questions:

- Can you think of other examples of social media research?
- Have you participated in social media research?
- What are other purposes of conducting social media research?

2. How can we conduct social media research ethically?

Given the pervasiveness of social media and data on SNS, it has become much easier to conduct research on social media. However, this means that there may be less checks and balances when it comes to conducting research ethically. Traditional means of ensuring that research is ethical may not be applicable to the digital environment. For this part of the tutorial, we will discuss the challenges of conducting research on social media.

Questions:

- To what extent do you think conducting ethical research from social media may be different to ethical research more generally?
- Given that formal ethics applications and consent procedures may not work for social media research, what do you think are possible solutions for conducting such research?
- Do you think conducting ethical research can help ensure that social network data is gathered more ethically?

3. Guidance for conducting social media research

For the final part of the tutorial, we will look at guidance for conducting social media research. We will read excerpts from the University's social media research ethical guidance as well as external policies that support ethical research.

Questions:

- What do you think about the guidance and policies that we read? Are they useful for researchers or for participants?
- What other things do you think should be included in social media research ethical guidance and policies?
- Do you think guidance and policies are enough to ensure that social media research is conducted ethically?

If you are interested in the content of the tutorial, please find a few resources below:

- "Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach" Carole Cadwalladr and Emma Graham-Harrison, The Guardian.
- "Internet Research Ethics for the Social Age: New Challenges, Cases, and Contexts" edited by Michael Zimmer and Katharina Kinder-Kurlanda.
- University social media research ethical guidance.

Appendix D

Adapted IUIPC questions

The following statements are IUIPC questions adapted for online learning included in the final survey of the study (Chapter 6). The statements were presented in Likert matrices with five responses available, ranging from strongly disagree to strongly agree.

The following statements relate to privacy practices:

- Online learning platforms should disclose the way my personal data is collected, processed, and used.
- Universities should disclose the way my personal data is collected, processed, and used.
- It is very important to me that I am aware and knowledgeable about how my personal information will be used.

The following statements relate to control over your personal data:

- Users' online privacy is really a matter of users' right to exercise control over how their information is collected, used, and shared.
- I believe that online privacy is violated when control over how users' information is collected, used, and shared is lost.

The following statements relate to data collection:

- It bothers me when online learning platforms ask me for personal information.
- It bothers me when universities ask me for personal information.
- When online learning platforms ask me for personal information, I sometimes think twice before providing it.
- When universities ask me for personal information, I sometimes think twice before providing it.
- It bothers me to give personal information to so many online companies.

Appendix E

Ethics Approvals

Two of the experiments discussed in this thesis involved human participation and were thus scrutinised and approved by the University of St Andrews' Teaching and Research Ethics Committee (UTREC). Confirmation of approval for both of these experiments, discussed in Chapters 4 and 6 respectively, are included on the following pages.

University Teaching and Research Ethics Committee

18 February 2020

Dear Janis,

Thank you for submitting your ethical application, which was considered by the School of Computer Science Ethics Committee on Tuesday 18th February, where the following documents were reviewed:

1. Ethical Application Form
2. Participant Information Sheet
3. Participant Consent Form
4. Participant Debrief Form
5. Interview Questions

The School of Computer Science Ethics Committee has been delegated to act on behalf of the University Teaching and Research Ethics Committee (UTREC) and has granted this application ethical approval. The particulars relating to the approved project are as follows -

Approval Code:	CS14765	**Approved on:**	18.02.20	**Approval Expiry:**	18.02.2025
Project Title:	Identifying and Understanding Challenges in the Urban Commons Process				
Researcher(s):	Janis Wong				
Supervisor(s):	Dr Tristan Henderson and Professor Kirstie Ball				

Approval is awarded for five years. Projects which have not commenced within two years of approval must be re-submitted for review by your School Ethics Committee. If you are unable to complete your research within the five year approval period, you are required to write to your School Ethics Committee Convener to request a discretionary extension of no greater than 6 months or to re-apply if directed to do so, and you should inform your School Ethics Committee when your project reaches completion.

If you make any changes to the project outlined in your approved ethical application form, you should inform your supervisor and seek advice on the ethical implications of those changes from the School Ethics Convener who may advise you to complete and submit an ethical amendment form for review.

Any adverse incident which occurs during the course of conducting your research must be reported immediately to the School Ethics Committee who will advise you on the appropriate action to be taken.

Approval is given on the understanding that you conduct your research as outlined in your application and in compliance with UTREC Guidelines and Policies (http://www.st-andrews.ac.uk/utrec/guidelinespolicies/). You are also advised to ensure that you procure and handle your research data within the provisions of the Data Provision Act 1998 and in accordance with any conditions of funding incumbent upon you.

Yours sincerely

Wendy Boyter

School Ethics Committee Administrator

School of Computer Science Ethics Committee

12 April 2021

Dear Janis,

Thank you for submitting your ethical amendment application.

The School of Computer Science Ethics Committee has approved this ethical amendment application:

Original Approval Code:	CS15295	**Original Approval Date:**	18.02.2021
Amendment Approval Date:	12.04.2021	**Approval Expiry Date:**	18.02.2026
Project Title:	Online Learning as a Commons		
Researcher(s):	Janis Wong	**Supervisor/PI:**	Dr Tristan Henderson
School/Unit:	Computer Science		

[Delete if not applicable] The following supporting documents are also acknowledged and approved:

1. Ethical Amendment Application Form
2. Advertisement
3. Survey

This approval does not extend the originally granted approval period. If you require an extension to the approval period, you can write to your School Ethics Committee who may grant a discretionary extension of no greater than 6 months. For longer extensions, or for any further changes, you must submit an additional ethical amendment application. For all extensions, you should inform the School Ethics Committee when your study is complete.

You must report any serious adverse events, or significant changes not covered by this approval, related to this study immediately to the School Ethics Committee.

Approval is given on the following conditions:

- that you conduct your research in line with:
  - the details provided in your ethical amendment application (and the original ethical application where still relevant)
  - the University's Principles of Good Research Conduct
  - the conditions of any funding associated with your work
- that you obtain all applicable additional documents (see the 'additional documents' webpage for guidance) before research commences.

You should retain this approval letter with your study paperwork.

School of Computer Science Ethics Committee
Dr Juan Ye/Convenor, Jack Cole Building, North Haugh, St Andrews, Fife, KY16 9SX
T: 01334 463252 E: ethics-cs@st-andrews.ac.uk

University of St Andrews | FOUNDED 1413

School of Computer Science Ethics Committee

Yours sincerely,

Wendy Boyter

SEC Administrator

School of Computer Science Ethics Committee
Dr Juan Ye/Convenor, Jack Cole Building, North Haugh, St Andrews, Fife, KY16 9SX
T: 01334 463252 E: ethics-cs@st-andrews.ac.uk